HUMBOLDT-UNIVERSITÄT ZU BERLIN

Combining paleontological and neontological data to assess the extinction risk of amphibians

Die Kombination paläontologischer und neontologischer Daten zur Beurteilung des Aussterberisikos von Amphibien

Dissertation

zur Erlangung des Akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

eingereicht an der Lebenswissenschaftlichen Fakultät der
Humboldt Universität zu Berlin

von
M. Sc. Melanie Tietje

Präsidentin der Humboldt Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachter/innen

1. PD Dr. Mark-Oliver Rödel
2. Prof. Dr. Nadia Fröbisch
3. Associate Professor Erin E. Saupe, PhD
4. Prof. Dr. Liliane Rueß

Tag der mündlichen Prüfung: 18. Dezember 2018

Leibniz Association

MUSEUM FÜR NATURKUNDE BERLIN

Combining paleontological and neontological data to assess the extinction risk of amphibians

Abstract

The extinction risk of a species is not random, but rather shaped by several factors comprising geographical, environmental and morphological traits. Some of these traits have been incorporated in assessment procedures for the classification of extant species' extinction risk, such as the International Union for Conservation of Nature (IUCN) Red List. These assessments are an important tool for conservation purposes, as they direct the available resources to species that are most reliant on support. This is especially important for amphibians, which today represent the vertebrate taxon with the highest proportion of threatened species.

However, a large number of species lack an assessment for extinction risk. Also, additional verification of the general influence of incorporated traits on extinction risk is still needed, as real extinction events are difficult to detect on neontological time scales. The fossil record offers the opportunity to test the influence of certain traits on extinction risk as it provides an enormous archive of extinction events that already happened, together with some of the traits of the species that have gotten extinct.

In this thesis, I examine traits in amphibian species that contribute to the extinction risk of this increasingly endangered group and provide support for the importance of geographic range size on the extinction risk of species. Placed in the developing field of Conservation Paleobiology, the study concentrates on the connection between paleontological and neontological data and how this unique combination can add to the knowledge about traits that shaped the extinction risk of amphibian species. This is achieved by quantitatively investigating species traits, conserved in the amphibian fossil record, and combining these findings with results from the IUCN Red List and climate data.

I show in paper 1 that the habitat type, one of the factors likely influencing extinction risk today, showed a contrasting influence in the fossil record dataset. While for extant species flowing water bodies are a potential factor for an increased extinction risk, species from similar habitats in the fossil record showed a decreased extinction risk. Alterations in the suitability of the habitat type might be a likely cause for this pattern, as anthropogenic influences are known to have disproportionally affected flowing waters.

The central aspect of this thesis is addressed in paper 2 and shows the prediction of the extinction risk of extant amphibian species with a model based entirely on paleontological data of extinct species. This model achieved high consensus between predicted duration and the extinction risk of species assessed via the IUCN Red List, showing that the combination of large scale paleontological and neontological data is possible. In this model, the geographic range size was supported as the most influential factor for extinction risk of species. The application of this paleo-fitted model on Data Deficient species, furthermore, shows the likely high extinction risk of this group. Hence this model is an additional option for identifying species which require protection, even for species with lacking data on population dynamics.

In paper 3, I used the novel connection of the amphibian fossil record and climate data derived from oxygen isotopes to investigate the temporal stability of the most influential trait for extinction risk; the geographic range size. The influence of this trait on extinction risk proved to be temporally stable, which is an important result as it supports the major role of the range size for extinction risk even over long time periods and under varying environmental conditions. Additionally, the observed minor changes in the connection between range size and duration seem to be influenced by global climatic patterns, especially by the meridional temperature gradient.

A closer collaboration between the fields of paleontology and species conservation has often been suggested, because of the similarity in goals. The present dissertation shows possible applications of the fossil record to current questions in conservation biology and shows how a combination of both fields contributes to the understanding of factors that influence the extinction risk of species.

Kurzfassung

Das Aussterberisiko einer Art ist nicht zufällig, sondern wird von mehreren Faktoren bestimmt, die geografische, ökologische und morphologische Merkmale umfassen. Einige dieser Merkmale sind Teil der Kriterien zur Einschätzung der Gefährdung einer Art, wie zum Beispiel in der Roten Liste der Internationalen Union zur Bewahrung der Natur und natürlicher Ressourcen (IUCN). Diese Beurteilungen sind ein wichtiges Werkzeug für den Artenschutz, da sie eine Verteilung der Maßnahmen auf die am stärksten gefährdeten Arten ermöglichen. Dies ist besonders wichtig für Amphibien, die Wirbeltiergruppe mit dem derzeit höchsten Anteil an bedrohten Arten.

Bei einem großen Teil der Arten fehlt jedoch eine Einschätzung des Aussterberisikos. Weiter mangelt es auch an einer endgültigen Verifizierung des Einflusses der genutzten Merkmale auf das Aussterberisiko, da Aussterbeereignisse auf neontologischen Zeitskalen schwer zu erkennen sind. Der Fossilbericht stellt ein enormes Archiv an bereits geschehenen Aussterbeereignissen dar, in dem auch verschiedene Merkmale der ausgestorbenen Arten erhalten sind. Damit bietet er die Möglichkeit den Einfluss bestimmter Merkmale auf die Gefährdung zu testen.

In der vorliegenden Arbeit untersuche ich Merkmale von Amphibienarten, die zum Aussterberisiko dieser zunehmend gefährdeten Gruppe beitragen und bestätige die Bedeutung der geographischen Reichweite für das Aussterberisiko. Die in dem sich aktuell entwickelnden Gebiet Conservation Paleobiology angesiedelte Arbeit konzentriert sich auf die Verbindung von paläontologischen und neontologischen Daten, und wie diese einzigartige Kombination dazu beitragen kann das Wissen über Aussterberisiko-beeinflussende Faktoren zu erweitern. Dies wird durch die quantitative Analyse verschiedener im Fossilbericht überlieferter Artmerkmale und der Kombination der daraus gewonnenen Erkenntnisse mit Ergebnissen der Roten Liste zu rezenten Arten und Klimadaten erreicht.

Ich zeige in Paper 1, dass der Einfluss des Habitattyps, ein vermutlich einflussreicher Faktor für das Aussterberisiko, im Fossilbericht einen gegenteiligen Einfluss zeigt als heute. Während für rezente Arten Fließgewässer ein potenzieller Faktor für ein erhöhtes Aussterberisiko sind, zeigen fossile Arten in vergleichbaren Habitaten ein verringertes Aussterberisiko. Änderungen in der Eignung des Habitats selbst könnten eine mögliche Ursache hierfür sein, da Fließgewässer unverhältnismäßig stark durch anthropogene Einflüsse beeinträchtigt werden.

Der zentrale Aspekt dieser Arbeit wird in Paper 2 behandelt und zeigt die Vorhersage des Aussterberisikos von rezenten Amphibienarten mit einem Modell, das ausschließlich auf paläontologischen Daten ausgestorbener Arten beruht. Dieses Modell erreicht einen hohen Konsens zwischen der vorhergesagten Dauer und der Einschätzung des Aussterberisikos durch die IUCN und zeigt damit, dass die Kombination von paläontologischen und neontologischen Daten möglich ist. In diesem Modell war die geographische Verbreitung der einflussreichste Faktor für das Aussterberisiko . Die Anwendung dieses mit Paläo-Daten erstellten Modells auf Datendefizitäre Arten zeigt die wahrscheinlich starke Gefährdung dieser Gruppe. Das Modell stellt damit eine zusätzliche Option zur Identifizierung bedrohter Arten dar, auch für Arten mit fehlenden Daten zur Populationsdynamik.

In Paper 3 habe ich mit der innovativen Verbindung des Amphibienfossilberichts und Klimadaten, abgeleitet aus Sauerstoffisotopen, die zeitliche Stabilität des Einflusses der geographischen Verbreitung auf die Gefährdung untersucht. Der Einfluss der Verbreitung auf die Gefährdung stellte sich hierbei als zeitlich stabil heraus und unterstreicht dabei die wichtige Rolle der geographischen Verbreitung für das Aussterberisiko auch über lange Zeiträume und unter unterschiedlichen Umweltbedingungen. Darüber hinaus scheinen die beobachteten geringfügigen Schwankungen zwischen Verbreitung und Gefährdung von globalen Klimamustern beeinflusst zu werden, insbesondere vom meridionalen Temperaturgradienten.

Eine engere Zusammenarbeit zwischen den Forschungszweigen Paläontologie und Artenschutz wurde wegen den ähnlichen Zielen oft angeregt. In meiner Arbeit zeige ich mögliche Anwendungen des Fossilberichts auf aktuellen Themen des Artenschutzes und wie eine Kombination beider Bereiche zum tieferen Verständnis von Gefährdungsfaktoren beitragen kann.

Preface and author's contributions

This thesis consists of an introductory part, which includes the background of the study and a section that recapitulates the most important findings from the papers and relates them to the original working hypothesis. The results of this thesis are presented in three scientific papers of which two are already published and the third is about to be submitted. The contributions of the co-authors to each paper are as listed below. From all three papers, the PhD candidate (M.T.) is the first author.

The papers are tackling the following themes:

Paper1
Tietje, M., Rödel, M.-O. (2017) Contradicting habitat type-extinction risk relationships between living and fossil amphibians, Royal Society Open Science, 4: 170051.
The first-author (M.T.) designed the paper, collected the data and carried out the analysis. Both authors contributed to the discussion and conception of the manuscript.

Paper 2
Tietje, M., Rödel, M.-O. (2018) Evaluating the predicted extinction risk of living amphibian species with the fossil record, Ecology Letters 21(8): 1135-1142
The first author (M.T.) performed data collection, modelling and analysis. Both authors contributed to the design of the study and writing of the manuscript.

Paper 3
Tietje, M., Schobben, M., Rödel, M.-O. (in prep.) Climate influences the effect of range size on extinction risk in the amphibian fossil record
This paper was designed and written by the first author (M.T.) with contributions from the co-authors. The first author performed research and analysed data, both co-authors contributed to the conception and discussion. M.S. provided expertise on data analysis.

Acknowledgements

Firstly, I would like to express my appreciation and thanks to my supervisor Mark-Oliver Rödel for his support and advice. Much appreciation and thanks goes out to the current and former members of the herpetology working group from the Museum für Naturkunde, in particular my close colleagues of which some have become dear friends who helped me in many phases of my research, and Schorsch. All the people mentioned beforehand are biologists, but much appreciation also goes out to those colleagues from the museums' paleontology department who gave valuable advice on my work and even let me be part in some of their own projects. As a biologist by training, working somewhere between biology and paleontology, I always juggled with the specifities of two research areas and always much appreciated help from both sides.

In addition, I would like to thank my co-authors Mark-Oliver Rödel and Martin Schobben for their contributions to the respective manuscripts and Brandon Kilbourne and Sami Assad for English spelling and grammar checks of the manuscripts.

Thank you to Nadia Fröbisch and Erin Saupe for kindly agreeing to review this thesis.

For the eternal struggle with R, the programming club founded by Johan Renaudie was of great help. Thank you also to all the scientists who participated in building and maintaining the Paleobiology Database and the people conducting the IUCN Red List assessments without whom this study would have been nearly impossible.

I would also like to mention the former and current doctoral student representatives of the museum who achieved a lot for our interdisciplinary education over the last years. It's been fun to be part of the team.

Thank you xkcd comics, The Friday Drinks, and Sanja for introducing me to reddit and therefore prolonging my thesis for about 2 months.

Thanks to my favorite paleontologist Richard for terrible puns and bringing my coffee addiction back to life.

The funding of my work took some unexpected turns, still I am grateful for the following funding sources: Elsa-Neumann for generously providing me with funding for 2.5 years, Wikimedia and the Stifterverband for accepting me as a fellow in their Open Science program, the Museum für Naturkunde for granting me funding for various conference trips, and the botany working group of the Humboldt Universität zu Berlin.

Last but not least. I want to thank my mum and dad for enabling me to concentrate on my studies in the first place. Many thanks to all friends who had to drink beer and listen to my worries over the last years. A very special thank you goes out to Martin. You have been with me for the most time of my thesis and have seen all the small and big catastrophes, and ups and downs, and not only kept me going, but also helped significantly improve my knowledge on geology, ggplots, snowboarding and bike mechanics.

Roses are Red
Violets are Blue
Unexpected {
In line 32

Contents

1 Introduction

1.1 Background

Amphibians are commonly considered to be the most endangered vertebrate taxon (Baillie et al., 2010, figure 1). Measures to help those endangered species to fight imminent extinction are needed, but to effectively allocate the restricted resources in conservation, a best possible identification of the most indigent species is crucial.

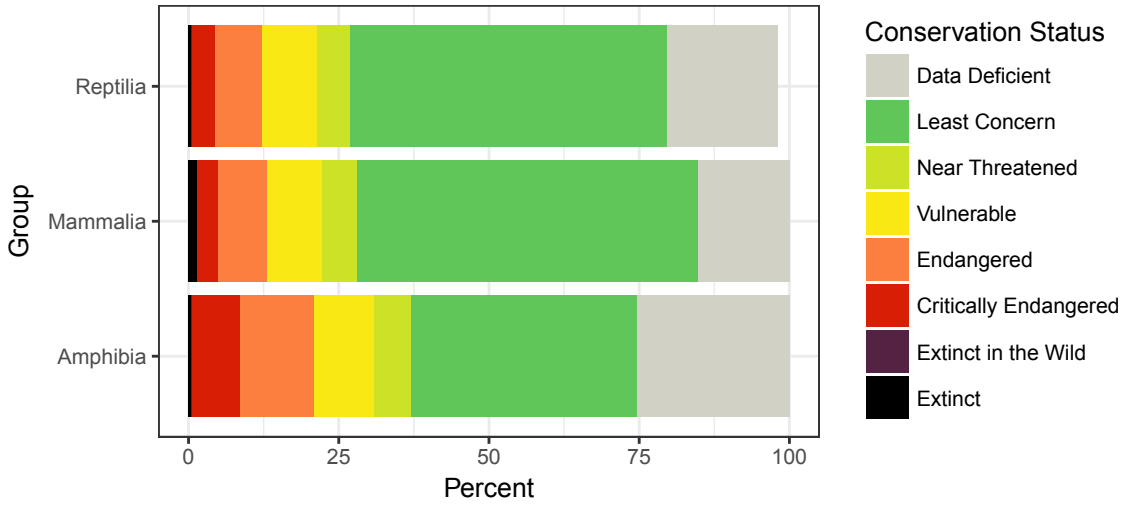

Figure 1 Proportion of the conservation status for the three terrestrial vertebrate groups amphibians, mammals and reptiles as assessed by the IUCN Red List. Amphibians show the highest proportion of Data Deficient species, the lowest proportion of Least Concern species, and also the highest proportions in the three groups classifying a species as threatened (Vulnerable, Endangered and Critically Endangered).

The way to estimate a species extinction risk has undergone several revisions since the earliest attempts in the 1950ies (Mace et al., 2008). These revisions included some editorial changes and clarifications, like the further clarification and purpose of the criteria and the inclusion of a section about data uncertainty and how to deal with it. Additionally, changes to the criteria itself were made, like the adjustment of quantitative thresholds for population sizes, or the additional differentiation between geographic range size measures to account for scaling problems between different taxa (Hilton-Taylor (Compiler), 2000). The current procedure to estimate extinction risk by the IUCN requires the application of five criteria; A) population size reduction, B) Geographic range, C) Small population size and decline, D) Very small or restricted population, E) Quantitative analysis. These criteria include information on the population size and its dynamics, the geographic range size and its dynamics, as well as other quantitative analysis methods (IUCN, 2012). This assessment procedure is based on the knowledge that extinction does not happen randomly, but that specific attributes influence the likelihood of species to become endangered and/or eventually extinct.

The search for these attributes is one of the major goals in conservation biology (Purvis et al., 2000). Common patterns in traits influencing extinction risk over a broader range of taxa and time periods are of special interest, as they yield the promise to more easily identify species that are likely to be especially prone to threat. Studies showed that while some traits like the geographic range size of species, the home range or population density show a consistent pattern in their influence on extinction risk across several taxa, other traits like the generation time or body size show differing results (Purvis et al., 2000).

Specifically, for extant amphibians it has been shown that the geographic range size has

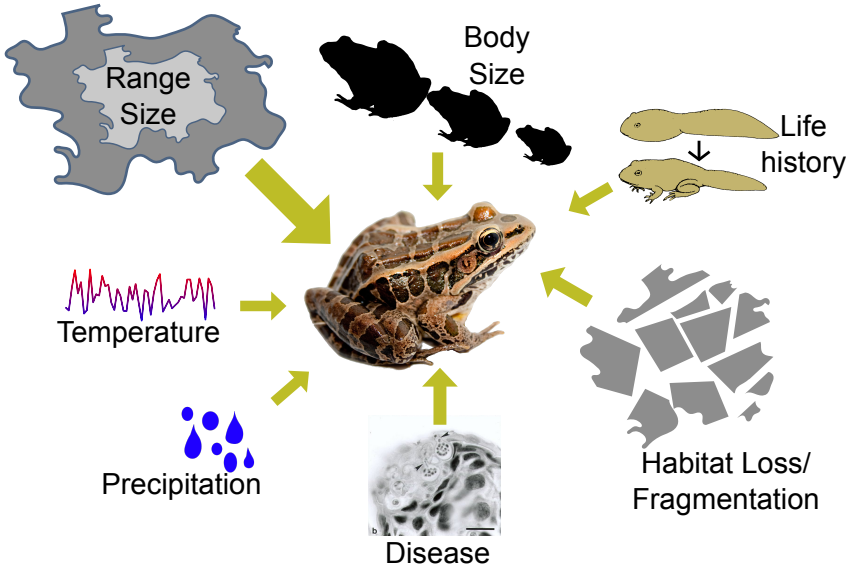

Figure 2 Major variables affecting amphibian species extinction risk, adapted from Sodhi et al. (2008). Source for the disease picture: By CSIRO, CC BY 3.0, `https://commons.wikimedia.org/w/index.php?curid=35445059`

a strong negative correlation with extinction risk. Other factors having an additional influence on extinction risk comprise the body size, the habitat and its environmental variables like precipitation and temperature, life history traits like the reproduction mode, or diseases like the amphibian chytrid fungus (Cooper et al., 2008; Sodhi et al., 2008; Hirschfeld and Rödel, 2017, figure 2). Still, geographic range size was in all studies the most important factor, which is likely explained by the ability of wide-ranging species to buffer local extinction events with other populations.

An individual assessment of each of the currently known 7890 amphibian species (AmphibiaWeb, 2018) is hardly possible, as it is also reflected by the number of currently assessed 6609 species, of which 1515 species are classified as Data Deficient, leaving effectively about one third of amphibian species without an estimation about their status. The lack of assessments has several reasons, among them financial restrictions, limited work-power and the possibly restricted accessibility of species in remote habitats (Purvis et al., 2000).

Although the IUCN Red List assessment criteria are thoroughly developed and have undergone repeated adjustments, the verification that the eventual extinction was caused, or at least influenced by, the traits that led to its assessment as threatened is difficult. Much of the difficulty stems from the restricted abilities to actually observe extinction events, which emerge from study periods being too short to cover the whole process, or from failing to meet the formal requirements to actually declare a species extinct. The IUCN Red List, as an example, requires extensive surveying of the complete potential habitat of an extinction candidate species, which also includes remote areas Purvis et al. (2000). As a result, there remains some uncertainty about whether or not the suspected traits eventually do lead to extinction. Therefore, it inevitably needs real extinction events for the verification of the influence of these traits on extinction risk (Mace et al., 1992).

The fossil record yields an extensive archive of species that eventually have gotten extinct. Together with these numerous extinction events, several traits of these species have been preserved as well. And like in conservation biology, the search for attributes that influence the extinction probability, or longevity, of species has been of interest in paleobiology, since at least the 1980ies, and especially since datasets suitable for macroecological studies have become more easily available (Harnik et al., 2012; McKinney, 1997; Sheehan and Hansen, 1986; Jablonski, 1986; LaBarbera, 1986). Pointing out the similar interests between these two fields of ecological

research, several studies have suggested and encouraged a closer connection of both research areas to benefit both (Dietl and Flessa, 2011; Fritz et al., 2013). These publications being opinion papers, so far still only one study has made actual use of the fossil record to apply the findings directly to extant species to learn about their extinction risk (Finnegan et al., 2015).

This thesis consists of 3 studies that are practical implementations of the called-for closer connection of paleobiology and conservation, in the evolving field of Conservation Paleobiology.

1.2 Working hypothesis

In this study I tested several aspects of the integration of paleontological and neontological data.

In a first step I address a major criticism towards paleontological data, the quality of the fossil record. As it is common knowledge, the fossil record is never complete and only depicts a small proportion of the species that have roamed the earth. Therefore a proper assessment of the quality and the potential biases of the amphibian fossil record lays the basis for all following analysis. I tested the quality of the fossil record by using four common techniques / metrics; the preservation potential of amphibians, the completeness of their fossil record, the reliability of short-lived species (single-interval taxa), and for preservation differences between habitat types to account for the unique life-cycle of amphibians and this potential bias on the fossil record. While testing for habitat preservation differences, I tested for the influence of the habitat type on the duration. Habitat type, although not included in the Red List procedures, is a factor that seems correlated with species extinction risk today. The comparison of the correlations of this factor with extinction risk today and in the deep time fossil record provides information about the temporal stability of its influence.

In a second step, I modeled the duration of extinct amphibian species using the collected trait information from databases and the literature to 1) identify the traits that are most important for extinction risk in amphibian species, 2) assess the applicability of a paleo-based model for the prediction of extant species' extinction risk. The identification of influential traits for extinction risk is important as it can serve as verification of the usage of these traits in extinction risk assessments of extant species. Additionally, the application of this model on extant species is an important test, as it bridges the deep time scale of the fossil record and the short-time scale of neontological ecology (like the IUCN Red List assessments).

As a third step, I conducted a detailed investigation of the temporal stability of the geographic range size and the possible influence of climate on the importance of this trait. Geographic range size turned out to be the most important factor for the duration of species in the model established in the second paper. Therefore, assessing the temporal stability of its influence on the extinction risk is important to estimate the universal applicability of this factor, especially in the light of changing environments and climate that we are currently facing. I expected the influence of the geographic range size on extinction risk to be temporally stable, given its ubiquitous appearance in studies on this matter in numerous taxonomic groups and over differing time scales. However, especially for amphibians as ectothermic organisms, climate variables like the meridional temperature gradient might have an influence on the geographic range size, and therefore indirectly on the extinction risk of a species.

1.3 Relevance of the study

The aim of this thesis is to increase our ability to identify endangered species, which ultimately aid in their conservation. Reasons to protect these species are rather simple: Amphibians play an integral role in the functioning of ecosystems as well as having an esthetic and economic value (Endangered Species Act of 1973). Their loss would therefore eventually affect human society as well. Hence the protection of amphibians is of high significance.

This study is as well the first of its kind that combines paleontological large-scale data with neontological data in vertebrates, a concept that has been suggested several years ago but not put into practice so far (Dietl and Flessa, 2011; Fritz et al., 2013). The results of this work proved promising for the developing field of conservation paleobiology.

In the first paper (section 5.1), I have shown that the anthropogenic habitat alteration seems to reverse the effect that the habitat type has on extinction risk. Knowing that this relationship was different in the past potentially shows the vast extent of anthropogenic influence on these habitats, marking them as aims of conservation efforts.

In the second paper (section 5.2), I provide validation for the vast influence of geographic range size on the extinction risk using extinction events in the fossil record. I have shown that data on extinction events from the fossil record can be used to provide additional knowledge on extant species, besides differing time scales.

The importance and moreover stability of the range size gets further evaluated in the third paper (section 5.3), where I show the temporal stability of the influence of the range size on extinction risk, and moreover investigate the connection of this influence with long-term climatic trends derived from oxygen isotope data.

2 Synthesis

2.1 The amphibian fossil record

The fossil record of amphibians comprises fossils from the last 330 million years, reaching back to the Late Carboniferous. As for most vertebrate taxa, fossil amphibians are usually subjects to morphological studies, but rarely to quantitative ones. Reasons for this are likely the lower quantities in which amphibian, and vertebrate fossils in general, are made available. However, exceptional preservation examples yield deep insights into the past ecology of amphibians, like exceptionally preserved tadpoles from the Miocene in Spain (McNamara et al., 2009), including soft-tissue, or the finding of middle Jurassic tadpoles in China (Yuan et al., 2003). Especially the mostly bi-phasic life cycle of amphibians is a big advantage over other terrestrial vertebrate taxa, as part of the life (juveniles as well as adults) needs to be spent in habitats with preferential preservation; lake and stream deposits (Schoch, 2014). The information conserved within these fossils ranges from geographical information like geographic distribution of a species over environmental information to morphological traits like the body size and even behavioral traits (see figure 3 for an example of a well preserved *Triadobatrachus massinoti*, that has been used to study locomotion in this stem salientian).

Analyzing the data quality of the amphibian fossil record in paper 1 set up the framework for all of my following work based on the fossil data. Using the common completeness metrics, I confirmed that among vertebrate taxa, the fossil record of amphibians is one of the most incomplete (figure 1 paper 1), agreeing with the work of Fara and Benton (2000). Benton (2014) compared the fossil record of land vertebrates in terms of completeness with the echinoderm one and claims this record to be usually considered as "good".

However, the applied completeness metrics mainly aim at estimating taxonomic completeness to assess the quality of diversity measures. The aim of my study was to collect a variety of different trait combinations in species rather than estimating diversity indices. Acknowledging this, I emphasize the lacking evidence for strong biases into a certain direction in the fossil record, especially for the stratigraphic duration of species, which serves as the fossil equivalent of extinction risk thoughout all papers in this work. Following an approach suggested by Fitzgerald and Carlson (2006), no evidence was found for the geological stage level being too rough for an analysis on the species level, therefore artificially causing single-interval species, creating a severe bias in the response variable for the model in paper 2. I acknowledged the remaining uncertainty by fitting the prediction model in paper 2 on a reduced dataset from which all single-interval

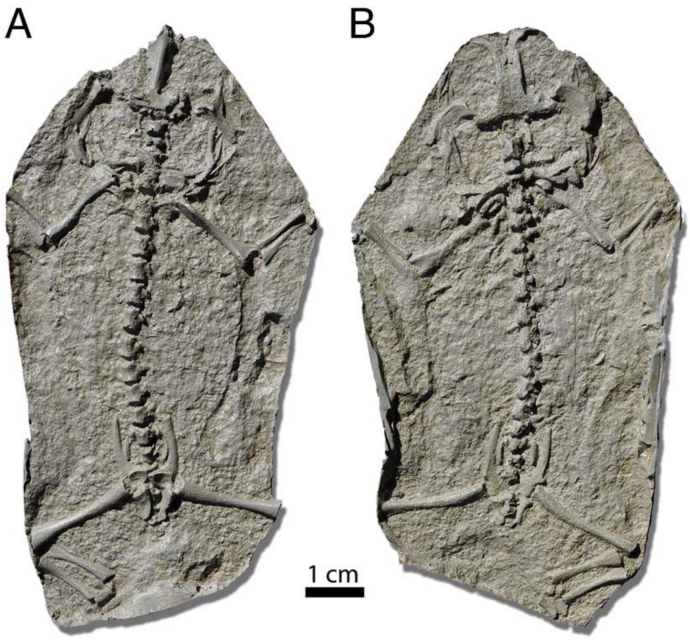

Figure 3 High-resolution cast of *Triadobatrachus massinoti*; Early Triassic of Madagascar, as an example of good preservation that allows for studies on locomotion. A) Ventral, B) dorsal view. From Lires et al. (2016)

species were removed, as it is common practice in most quantitative analysis. While showing a larger uncertainty in the prediction, indicated by a higher Root Mean Squared Error (RMSE, supplement paper 2, table S5), the explanatory power of the model remained comparable. It has to be mentioned that, in contrast to paper 2, excluding single-interval species from the analysis in paper 1 leads to the loss of the formerly observed difference in duration between habitat types. The reason is likely the relatively weak influence of the habitat type on the duration in comparison to the geographic range size, the dominating variable of the prediction model in paper 2.

Testing for potential bias from Lagerstätten and monographic effects turned out negative as well, and a specimens completeness metric (Benton, 2008) showed as expected a minor influence of taxonomy on the preservation, likely due to morphological differences (heavy skulls of Temnospondyli). These differences disappeared when testing Lissamphibia taxa only, so excluding the earliest amphibians Temnospondyli and Lepospondyli, which I acknowledged in paper 2 by running a model fitted exclusively on Lissamphibia taxa (supplement figure S15), yielding comparable results (but again higher RMSE due to lower sample size, see supplement paper 2, table S5).

Inaccuracies in the fossil record can never be ruled out and are most likely included in this dataset as well. The important question remaining is, however, if missing parts are equally distributed among taxa, environments, and time. From the combination of 1) lack of evidence for strong specific biases, and 2) closely matching results when running the following analysis on subsets of the data to exclude potential biases, I concluded that the fossil record would be sufficient for the anticipated work.

2.2 The changing habitat influence

For living amphibians today, habitat type and extinction risk show the interesting correlation of extinction risk being potentially higher in flowing water (lotic) environments (Stuart et al., 2008). The IUCN Red List assessment results show that 36% of species from lotic habitats

are categorized as threatened (or even extinct), whereas it is only 15% endangered species in lentic habitats. One idea to explain this pattern includes the increased likelihood that animals are being exposed to diseases in flowing water bodies. However, the exact mechanism remains vague to this point, which led to the idea of paper 1, which investigates the habitat influence on duration of extinct amphibian species.

I detected increased extinction risk in fossil amphibian species from low-energy water habitats, which is in contrast to what is been observed today (figure 4).

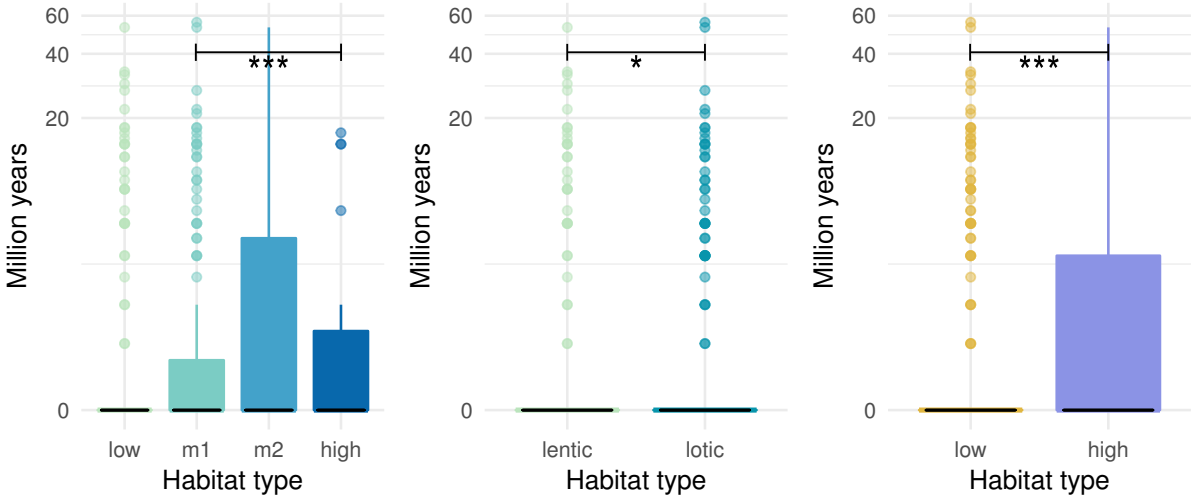

Figure 4 Modified from figure 3, paper 1. Durations of amphibian species in different environments. Species were grouped into four basic (left) and two broader environmental categories (mid and right). Black lines indicate the median and colored areas the range between first and third quartiles. Significant differences are indicated by one, two and three asterisks indicating p-values smaller than 0.05, 0.01 and 0.001 from Kruskal-Wallis rank sum test, respectively (for exact values compare electronic supplementary material, table S6).

In the discussion of paper 1, I argue that this apparent change in habitat influence on extinction risk might be explained by either changes in the habitat demands of amphibians, or changes in the habitat itself. Changes in amphibians life histories comprise the reduction in mean body size, with only single species maintaining impressive body sizes like the giant salamander *Andrias*, or the goliath frog (*Conraua goliath*). Larger body sizes were much more common in Temnospondyli, who have been characterized as "medium-sized to large aquatic predators" by Schoch and Milner (2000). With changing body sizes, habitat requirements have likely changed as well, however excluding the Temnospondyli from the analysis yielded similar results.

I concluded that the most likely course for this influence shift might be the varyingly strong alteration of the habitat suitability by human influence. Formerly natural river systems are being altered since earliest human settlements. Influences include habitat fractionation, the spread of exotic species, logging activities which affect also the macroinvertebrate communities, and a general disturbance of the habitat via mainly agricultural activities (Gillespie and Hero, 1999; Stone and Wallace, 1998; Willson and Doreas, 2003; Allan and Flecker, 1993; Everard and Moggridge, 2012). While these factors act to some extent on lentic water bodies as well, it might be the combination of more intensive human usage of flowing water bodies on one side, and a more generally higher vulnerability of species from lotic habitats in comparison to lentic species, which always had to deal with habitat alterations, like for example desiccation during the reproductive periods.

This unexpected discovery in paper 1 had two consequences for the following analysis. For once, the habitat type got excluded from the model that predicts the extinction risk of living species in the paper 2. A variable that seems to change its influence over time would not be of any help when trying to model the extinction risk of living species, as it assumes differing

conditions. As the reason for this change is hard to pin down, the exclusion of this variable seemed the safest thing to do.

Additionally, a closer investigation of the most important trait from the extinction risk prediction model, the geographic range size, was performed in paper 3. This analysis had the emphasis on temporal patterns in the trait itself and also in the influence strength of this trait.

2.3 Modelling extinction risk

Besides the long history of identifying traits that influence extinction risk, both in conservation and paleontology, the next step to actually model extinction risk for living species has been rarely made. Notable exceptions include for example the study by Howard and Bickford (2014), which uses a Random Forest classification model to predict the conservation status of data deficient amphibian species from geographical and life history traits. The model was fitted on available data from the IUCN Red List and applied to the data deficient species, with a success rate of 74% (based on the model validation). Like in Sodhi et al. (2008), the geographic distribution was the most influential factor in this model. As I point out in paper 2 however, the validation of current assessments is still of high importance and needs real extinction events. Therefore, the model by Howard and Bickford (2014) actually predicts the future assessment status of the species when more data will be available, which does not necessarily comply with real extinction risk.

A second study, which I partly closely followed methodologically, was published by Finnegan et al. (2015). They used data from the fossil record to fit a Generalized Boosted Regression Model (GBM) for predicting the intrinsic risk of modern genera. They consider their results as paleontological baselines for extinction risk that would be true for genera without any human influence. While the setup of this study is similar to paper 2 of my thesis, it comprises exclusively marine taxa, mostly invertebrates and marine mammals and sharks. The study does not compare the predictions with current IUCN Red List assessments of species from the included genera; however it compares the predicted extinction risks geographically with hotspots of human activity.

The best GBM that I constructed in paper 2 connects the stratigraphic range (duration) with several traits from fossil species and provided explanation for about 27% of the variation in the duration of species.

Geographic range size was, unsurprisingly, the most influential variable in this model (see table 1 paper 2), which agrees with plenty of studies from paleontological and neontological ecological research, as outlined in the introduction. The application of this model on the extant species dataset lead to predicted durations, which were compared with their IUCN Red List assessment status (figure 5). To exclude the possibility that this pattern is an artifact of the underlying model, caused for example by collinearity patterns in the input variables, I built an additional model with similar parameters, but fitted on a randomized version of the fossil dataset (Null model). This model performed much worse in explaining the duration of species and did not lead to any reasonable pattern when applied to the extant species (figure 5). Additionally, it needs to be pointed out that this prediction pattern remains mostly stable when changing the model parameters, the model type, or even the composition of the fossil data it was fitted with, for example by excluding single-interval species, or the early amphibian stem-taxa Temnospondyli and Lepospondyli (figure S20).

The modeling results agree with the predictions by Howard and Bickford (2014), who found that Data Deficient species should be on average at higher risk than most assessed species. Finding this pattern being suggested by both models from extant and extinct species emphasizes the urgency of proper assessments of these species.

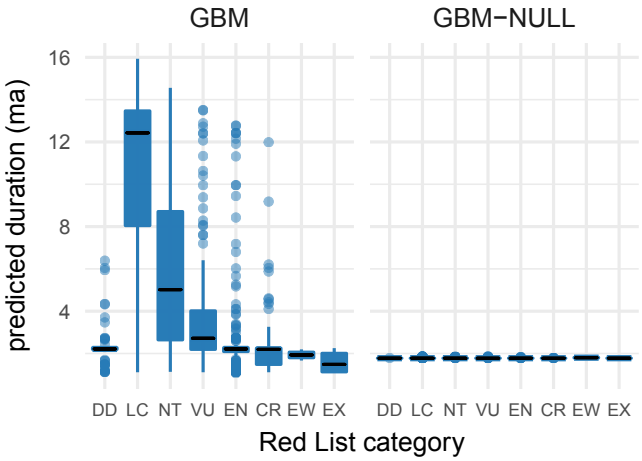

Figure 5 Predicted durations in million years for living amphibian species, based on the GBM models fitted to extinct species, figure taken from paper 2: Tietje and Rödel (2018). Left: final GBM, right: Null model. Model parameters were n trees = 200, interaction depth = 1, shrinkage = 0.01 and minimum number of observations in the trees terminal nodes = 5. Predicted durations are grouped by their species' IUCN Red List assessment status. GBM = Generalized boosted model, GBM-NULL = Null model on randomized data. Red List categories (with number of species) are Data Deficient (66), Least Concern (740), Near Threatened (98), Vulnerable (151), Endangered (192), Critically Endangered (128), Extinct in the Wild (2), Extinct (5).

2.4 Temporal stability

The temporal stability of trait influence on extinction risk is a critical factor that needs to be addressed when combining datasets from different times and also time scales. Larger instabilities would question the general applicability of large-scale ecological patterns on current questions in, for example, conservation. Moreover, I found in paper 1 that the influence of the habitat on extinction risk seems to have changed over time, providing evidence that not all patterns observed in the fossil are readily applicable for living species.

Geographic range size was the most important factor influencing extinction risk in paper 2, in accordance with a vast number of publications from paleontology and neontological ecology. Hence, if there are temporal variations in the influence strength of traits, it would be most easily detectable in the range size.

To test the temporal stability of the influence of range size on extinction risk, I divided the dataset into geological periods and tested the correlation strength of geographic range size and duration in each period (figure 6). All results where within two times standard deviation of the Null model, showing that the influence strength does vary slightly over time, but none of the variations are significant. This supports the applicability of the results from paper 2 by adding evidence for the temporal stability of the influence of geographic range size.

As climatic conditions are known to be influential for geographic range sizes, maybe even more so for amphibian species as ectotherms, I added climate data on the global mean temperature and the global meridional temperature gradient to the analysis, both derived from available oxygen isotope data (Veizer and Prokoph, 2015). While the mean global temperature did not seem to be of any effect, the temperature gradient was correlated with both geographic range size, and the correlation strength of range size with duration (figure 2 and S4, paper 3). The latter finding would potentially show that the influence strength of geographic range size on extinction risk varies with the meridional temperature gradient; however few of the linear models fitted on the correlation data are significant due to small sample size (table 1, paper 3).

A large temperature gradient likely corresponds to smaller potentially suitable habitats, and

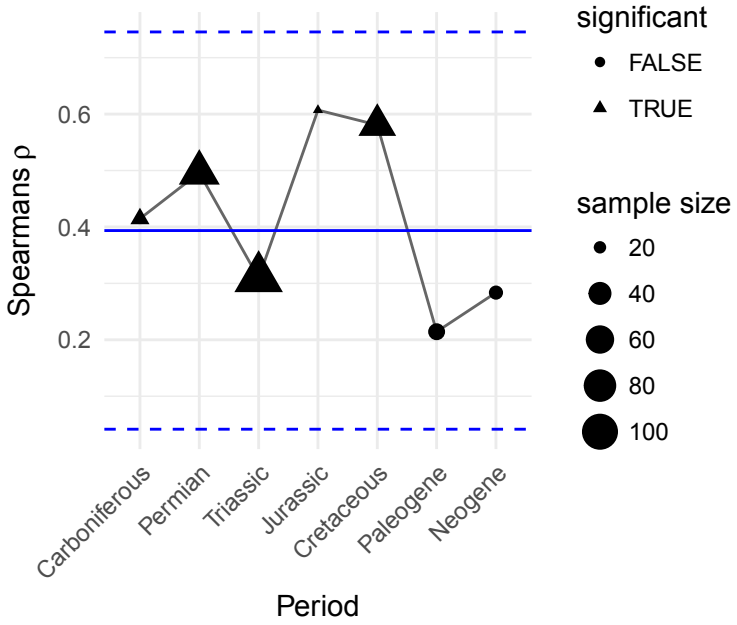

Figure 6 Correlation of geographic range size and duration of species in different geologic periods, taken from paper 3. Correlation was assessed using Spearmans rank correlation. The number of species per period visualized as the size of the dots, significance of the correlation is visualized as shape of the data points, with p-values smaller than 0.05 being visualized as triangles. The solid blue line depicts the overall mean correlation, the dashed lines are the two-times standard deviation of correlation distribution gained from repeated calculation of the correlation from a random subset of the data.

it correlates with lower importance of the range size for extinction risk. A possible explanation might lie in stronger competition in smaller habitats, making competitive abilities more important for survival than a large spread. Further studies on this topic are required to aid to a better understanding of these entangled variables and the potential influence of climate on them.

3 Conclusions and outlook

The findings of this thesis provide new insights into the possibilities and challenges of combining paleontological and neontological data. I have shown that not all traits seem to be temporally stable (paper 1), while the geographic range size is stable, with minor variations being likely explained by climatic influence (paper 3). Additionally, the geographic range size was confirmed as the most important factor for extinction risk in amphibians (paper 2). Both results are encouraging that the combination of fossil and neontological data is possible, even with fossil records of taxa usually considered as rather incomplete.

Paper 3 shows the combination of climate data with paleobiological data from the fossil record, a rather new combination and in this extent a slightly rough analysis, given the time scales and the sample sizes from the fossil record for each time unit. However, there seems to be a very interesting relationship between range sizes and the meridional temperature gradient, which are worth to be investigated in more detail in the future, maybe also across different taxa, given the universal influence of geographic range size. The incorporation of geochemical data, providing proxies for environmental changes, and paleontological data seems to hold a valuable approach in the future.

4 Bibliography

Allan, J. D. and Flecker, A. S. 1993. Biodiversity conservation in running waters. *BioScience*, 43(1): 32–43.

AmphibiaWeb. University of California, Berkeley, CA, USA. Available at: https://amphibiaweb.org/. Last accessed 29 June 2018., 2018.

Baillie, J. E. M., Griffiths, J., Turvey, S. T., Loh, J., and Collen, B. *Evolution Lost: Status and Trends of the World's Vertebrates*. Zoological Society of London, 2010.

Benton, M. *Vertebrate palaeontology*. Blackwell Publishing, Oxford, UK, third edition, 2014.

Benton, M. J. 2008. How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology*, 34(4): 516–533.

Cooper, N., Bielby, J., Thomas, G. H., and Purvis, A. 2008. Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography*, 17: 211–221.

Dietl, G. P. and Flessa, K. W. 2011. Conservation paleobiology: putting the dead to work. *Trends in Ecology and Evolution*, 26(1): 30–37.

Everard, M. and Moggridge, H. L. 2012. Rediscovering the value of urban rivers. *Urban Ecosystems*, 15: 293–314.

Fara, E. and Benton, M. J. 2000. The fossil record of Cretaceous tetrapods. *Palaios*, 15(2): 161–165.

Finnegan, S., Anderson, S. C., Harnik, P. G., Simpson, C., Tittensor, D. P., Byrnes, J. E., Finkel, Z. V., Lindberg, D. R., Liow, L. H., Lockwood, R., Lotze, H. K., Mcclain, C. R., McGuire, J. L., O'Dea, A., and Pandolfi, J. M. 2015. Paleontological baselines for evaluating extinction risk in the modern oceans. *Science*, 348(6234): 567–670.

Fitzgerald, P. C. and Carlson, S. J. 2006. Examining the latitudinal diversity gradient in Paleozoic terebratulide brachiopods: should singleton data be removed? *Paleobiology*, 32(3): 367–386.

Fritz, S. A., Schnitzler, J., Eronen, J. T., Hof, C., Böhning-Gaese, K., and Graham, C. H. 2013. Diversity in time and space: wanted dead and alive. *Trends in Ecology & Evolution*, 28(9): 509–516.

Gillespie, G. and Hero, J.-M. Potential impacts of introduced fish and fish translocations on Australian amphibians. In Campbell, A., editor, *Declines and disappearances of Australian frogs*, pages 131–144. Canberra, ACT: Biodiversity Group, Environment Australia, 1999.

Harnik, P. G., Lotze, H. K., Anderson, S. C., Finkel, Z. V., Finnegan, S., Lindberg, D. R., Liow, L. H., Lockwood, R., McClain, C. R., McGuire, J. L., O'Dea, A., Pandolfi, J. M., Simpson, C., and Tittensor, D. P. 2012. Extinctions in ancient and modern seas. *Trends in Ecology & Evolution*, 27(11): 608–17.

Hilton-Taylor (Compiler), C. *2000 IUCN Red List of Threatened Species*. IUCN, Gland, Switzerland and Cambridge, UK, 2000.

Hirschfeld, M. and Rödel, M.-O. 2017. What makes a successful species? Traits facilitating survival in altered tropical forests. *BMC Ecology*, 17(1): 25.

Howard, S. D. and Bickford, D. P. 2014. Amphibians over the edge: silent extinction risk of data deficient species. *Diversity and Distributions*, 20(7): 837–846.

IUCN. *IUCN Red List Categories and Criteria: Version 3.1.* Gland, Switzerland and Cambridge, UK: IUCN. iv + 32pp, second edition, 2012.

Jablonski, D. 1986. Background and mass extinctions: the alternation of macroevolutionary regimes. *Science*, 231(4734): 129–133.

LaBarbera, M. The evolution and ecology of body size. In *Patterns and Processes in the History of Life*, pages 69–98. Springer, 1986.

Lires, A. I., Soto, I. M., and Gómez, R. O. 2016. Walk before you jump: new insights on early frog locomotion from the oldest known salientian. *Paleobiology*, 42(4): 1–12.

Mace, G., Collar, N., Cooke, J., Gaston, K., Ginsberg, J., Leader Williams, N., Maunder, M., and Milner-Gulland, E. 1992. The development of new criteria for listing species on the IUCN Red List. *Species*, 19: 16–22.

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akçakaya, H. R., Leader-Williams, N., Milner-Gulland, E. J., and Stuart, S. N. 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, 22(6): 1424–1442.

McKinney, M. L. 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annual Review of Ecology and Systematics*, 28: 495–516.

McNamara, M. E., Orr, P. J., Kearns, S. L., Alcalá, L., Anadón, P., and Peñalver-Mollá, E. 2009. Exceptionally preserved tadpoles from the Miocene of Libros, Spain: ecomorphological reconstruction and the impact of ontogeny upon taphonomy. *Lethaia*, 43(3).

Purvis, A., Jones, K. E., and Mace, G. M. 2000. Extinction. *BioEssays*, 22: 1123–1133.

Schoch, R. R. *Amphibian evolution: the life of early land vertebrates.* Number April 2012. Wiley-Blackwell, UK, 2014.

Schoch, R. R. and Milner, A. R. *Encyclopedia of Paleoherpetology 3B: Stereospondyli.* Verlag Dr. Friedrich Pfeil, München, Germany, 2000.

Sheehan, P. M. and Hansen, T. A. 1986. Detritus feeding as a buffer to extinction at the end of the Cretaceous. *Geology*, 14(10): 868–870.

Sodhi, N. S., Bickford, D., Diesmos, A. C., Lee, T. M., Koh, L. P., Brook, B. W., Sekercioglu, C. H., and Bradshaw, C. J. A. 2008. Measuring the meltdown: drivers of global amphibian extinction and decline. *PLoS ONE*, 3(2): e1636.

Stone, M. K. and Wallace, J. B. 1998. Long-term recovery of a mountain stream from clearcut logging: the effects of forest succession on benthic invertebrate community structure. *Freshwater Biology*, 39(1): 151–169.

Stuart, S., Hoffmann, M., Chanson, J., Cox, N., Berridge, R., Ramani, P., and Young, B., editors. *Threatened Amphibians of the World.* Lynx Edicions, Barcelona, Spain; IUCN, Gland, Switzerland; and Conservation International, Arlington, Virginia, USA, 2008.

Tietje, M. and Rödel, M.-O. 2018. Evaluating the predicted extinction risk of living amphibian species with the fossil record. *Ecology Letters*, 21(8): 1135–1142.

Veizer, J. and Prokoph, A. 2015. Temperatures and oxygen isotopic composition of Phanerozoic oceans. *Earth-Science Reviews*, 146: 92–104.

Willson, J. D. and Doreas, M. E. 2003. Effects of habitat disturbance on stream salamanders: implications for buffer zones and watershed management. *Conservation Biology*, 17(3): 763–771.

Yuan, C., Zhang, H., Li, M., and Ji, X. 2003. Discovery of a middle Jurassic fossil tadpole from Daohugou region, Ningcheng, Inner Mongolia, China. *Acta Geologica Sinica*, 78(2): 145–148.

5 Papers

Paper 1

Contradicting habitat type-extinction risk relationships between living and fossil amphibians

Melanie Tietje, Mark-Oliver Rödel, Contradicting habitat type-extinction risk relationships between living and fossil amphibians. R. Soc. open sci. 2017 4 170051; DOI: 10.1098/rsos.170051. Published 10 May 2017

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Tietje M, Rödel M-O. 2017 Contradicting habitat type-extinction risk relationships between living and fossil amphibians. *R. Soc. open sci.* **4**: 170051. http://dx.doi.org/10.1098/rsos.170051

Received: 17 January 2017
Accepted: 11 April 2017

Subject Category:
Biology (whole organism)

Subject Areas:
ecology/palaeontology

Keywords:
Amphibia, fossil record, extinction risk, habitat trait, Anthropocene

Author for correspondence:
Melanie Tietje
e-mail: melanie.tietje@mfn-berlin.de

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3757235.

THE ROYAL SOCIETY PUBLISHING

Contradicting habitat type-extinction risk relationships between living and fossil amphibians

Melanie Tietje[1] and Mark-Oliver Rödel[1,2]

[1]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstraße 43, 10115 Berlin, Germany
[2]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany

MT, 0000-0003-1157-2963

Trait analysis has become a crucial tool for assessing the extinction risk of species. While some extinction risk-trait relationships have been often identical between different living taxa, a temporal comparison of fossil taxa with related current taxa was rarely considered. However, we argue that it is important to know if extinction risk-trait relations are constant or changing over time. Herein we investigated the influence of habitat type on the persistence length of amphibian species. Living amphibians are regarded as the most threatened group of terrestrial vertebrates and thus of high interest to conservationists. Species from different habitat types show differences in extinction risk, i.e. species depending on flowing waters being more threatened than those breeding in stagnant sites. After assessing the quality of the available amphibian fossil data, we show that today's habitat type-extinction risk relationship is reversed compared to fossil amphibians, former taxa persisting longer when living in rivers and streams, thus suggesting a change of effect direction of this trait. Neither differences between amphibian orders nor environmentally caused preservation effects could explain this pattern. We argue this change to be most likely a result of anthropogenic influence, which turned a once favourable strategy into a disadvantage.

1. Introduction

With an increasing number of species being potentially threatened by anthropogenic environmental changes, scientists have been searching for new insights into how to determine the extinction risk of species [1–3]. Using traits—biotic and spatial characteristics of a species—has become a common method to estimate this risk [4,5]. The International Union for the Conservation of Nature

(IUCN) for example uses a set of trait categories to assess the risk of species for their Red List, which is a prime guide for conservation strategies [6].

When traits for a large variety of species are applied toward predictive conservation, it would be important to know that the effect of a certain trait is actually stable, or at least influences the extinction risk of different species in the same direction. Hence, it is of importance to test whether the targeted traits are valid in extinction risk assessments across different species and over longer time scales. Such validation would make the assessment process easier and comparable across a larger number of different taxa. For instance a widely applied trait is species' geographical range size [7]. Various examples from living and fossil species revealed geographical range size as a good proxy for extinction risk [8–10].

However, it has also been shown that some traits, like body size and life history, do not necessarily have the same effect size or even direction for differing taxa [8,11,12]. Given that differences in the response of certain traits are possible between taxa that exist today, the validity of traits over longer time scales should be of equal importance as the current state of species is always a product of their evolutionary past, and traits naturally evolve together with their species. Assessing the trait-extinction risk relationship in the past is possible by using the fossil record of species, which provides a great archive of multitudinous species with different trait combinations over various timespans [13,14]. Not all, but some of their traits, i.e. morphological ones, have been preserved over long time periods, and these can be combined with the species' longevity (duration). Thus, we can follow species from (an approximate) first to last appearance in the fossil record and are able to test for a correlation between certain traits and species' durations, consequently drawing inferences about the influence of traits on the extinction risk.

In this study, we intended to assess the impact of habitat traits on amphibian survival in the past. We chose to use amphibians for this task for several reasons: (i) amphibians are an old taxonomic class, reaching 370 million years back into the late Devonian [15,16]. This increases the chance to gather fossil species over a broad time scale; (ii) most amphibians have aquatic ontogenetic stages [17], which makes them dependent on bodies of water and should be beneficial for fossil preservation; (iii) current amphibian species are almost completely assessed for their conservation status and comprise the highest proportion of endangered species among terrestrial vertebrates [18,19] and thus are of high interest in current conservation efforts; and (iv) they exhibit a distinct difference in proportion of endangered species between lotic (flowing) and lentic (stagnant) freshwater habitats, lotic species being seemingly more threatened [20]. This habitat information is preserved in the fossil record.

We herein investigated whether an increased extinction risk in lotic amphibian species can be as well inferred from the fossil record. Based on data from two databases and further literature, we assessed the duration of extinct species in different habitat types. We tested for phylogenetic influence as well as for influence of habitat-dependent preservation bias. Before analysing our data, we evaluated the quality of the amphibian fossil record by applying several preservation and completeness metrics.

2. Methods

2.1. Data sources

The fossil amphibian data were collected from the Paleobiology Database (PbDb) on 15th February 2016 using http://fossilworks.org [21] and from the FosFARbase on 18th July 2016 using http://www.wahre-staerke.com [22], collecting all available amphibian data with no restrictions on geological time. Data downloaded included lithology, location and time (geological stage) of the fossil occurrences. Lithology data was solely available from the PbDb and complemented by literature research where possible (see supplement for lithology data with references). Downloads followed keyword searches with 'Allocaudata, Amphibia, Anura, Caudata, Gymnophiona, Temnospondyli, Urodela', and thus comprised amphibian orders including their stem groups (Salientia, Urodela and Parabatrachia), Lepospondyli, and the group of temnospondyl amphibians. Caudata and Urodela were used as interchangeable terms owing to differing taxonomic opinions of the two databases. By including Lepospondyli and Temnospondyli, we account for both the temnospondyl and lepospondyl origin hypothesis of modern amphibians [23]. Our exact usage of taxonomic terms is explained in the electronic supplementary material, table S1. Extinct species' names follow the taxonomy used in the PbDb. Taxonomy from fosFARbase was adapted to the names used in PbDb (electronic supplementary material, table S2). Duplicate records were manually excluded from the search results as were all species or genera marked with 'aff.', 'cf.' or '?'. We identified extant species according to their presence in the 'Amphibian species of the World' database [24] and excluded those species from the fossil list of taxa.

The final dataset contained 1658 occurrences from 620 extinct species, covering the amphibian fossil record from the Carboniferous to the Holocene. In this dataset, 816 occurrences from 358 species had lithology information and thus were included in the analyses of habitat influence on species duration. For comparison, we collected taxon lists for extant and fossil mammals from the PbDb (10th February 2016).

2.2. Data quality and completeness

As, for obvious reasons, data from the fossil record will be always incomplete, it is of high importance to test datasets for potential preservation flaws [25,26]. As information on the quality of the amphibian fossil record is sparse, we first assessed quality and completeness of our data by checking four factors that could potentially bias our results: (i) preservation potential of amphibians; (ii) completeness of the fossil record; (iii) reliability of single-interval species; and (iv) preservation differences between habitats.

The preservation potential was assessed by estimating the proportion of living taxa with a fossil record and by calculating the preservation probability. The proportion of living taxa with a fossil record shows the basic potential of a group to be preserved by counting how many living taxa of that group are already preserved as fossils [27]. Another preservation probability estimate uses the range-frequency distribution of fossil taxa. Using equation (2.1), following Foote and Raup [28] and Foote and Sepkoski [29], we calculated the preservation probability for amphibian species, genera and families:

$$\text{preservation probability} = \frac{f_2^2}{f_1 * f_3}. \tag{2.1}$$

Species duration was measured as number of geological stages from first to last occurrence of the species. The frequencies f_1, f_2 and f_3 represent the proportion of taxa with durations of one, two and three geological stages. Lower ratios indicate more taxa with short ranges and therefore a lower preservation probability. Completeness of the fossil record was checked by applying the simple completeness metric (SCM, [30]). The SCM measures completeness of the fossil record based on the gaps in each taxon's record. It was calculated as the ratio of observed fossil occurrences to total inferred fossil occurrences ([31], equation (2.2)). As result, lower SCMs show fewer gaps in the fossil record and therefore a higher completeness. Calculations were done on species, genus and family levels:

$$\text{SCM} = \frac{\text{known record}}{\text{assumed record}}. \tag{2.2}$$

As the fossil data included a high percentage of single-interval species, simply removing them would have resulted in a huge loss of diversity and the loss of potentially very shortly lived species. We thus investigated the reliability of short durations in these species following Fitzgerald and Carlson [32]. We determined the proportion of single-interval species from lagerstätten as well as monographic effects in our data to test for their effect on the number of single-interval taxa. Lagerstätten provide an exceptional amount and/or quality of preserved fossils [33], increasing the probability to find rare species. Similarly a focused sampling on one temporal interval can cause an increase in single-interval taxa (monographic effect), as it increases the probability to find rare species. Both factors intensify sampling of one region or time frame and can lead to a higher proportion of single-interval taxa. Rare species detected this way would probably be overlooked with less intense sampling, which would result in these species being falsely detected as single-interval species. We assessed the proportion of single-interval taxa described in monographies (publications which covered more than 20 occurrences) and compared those to the proportion in the complete dataset. Different proportions of single-interval taxa would indicate that the data suffer from monographic effects. We estimated the influence of lagerstätten by comparing the amount of single-interval species from lagerstätten and the complete dataset using Pearson's χ^2 test for count data (significance level $\alpha = 0.05$). Besides lagerstätten and monographic effects, we also tested for correlation between geological stage durations and species richness, number and proportion of single-interval species per stage. A correlation could indicate that single-interval species were not necessarily short-lived, but a result of the temporal resolution of the rock record, meaning the variability of stage durations might influence the amount of single-interval species. Correlation analyses were done using Spearman's rank correlation.

Finally, differences in preservation probability between habitats were assessed using the specimen completeness metric by Benton [34]. This metric assigns each occurrence increasing values from one to five, according to the conditions of the available specimens (isolated bones, one (nearly) complete skull, several skulls, one (nearly) complete skeleton, and several skeletons). High values indicate a

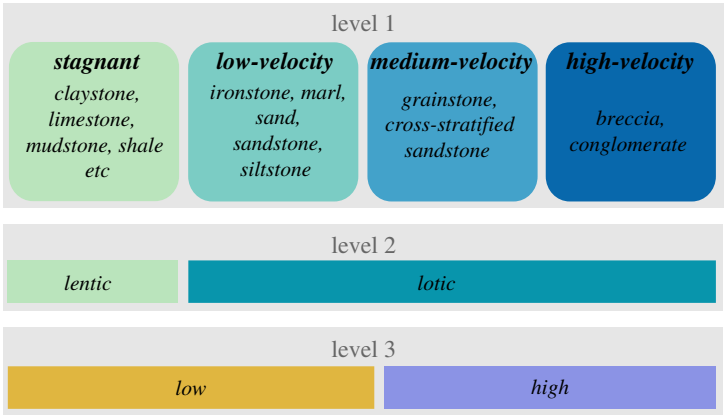

Figure 1. Lithologies were assigned to four habitat categories (*stagnant*, *low-velocity*, *medium-velocity* and *high-velocity*), reflecting an increase of water flow energy and thus a continuum from stagnant to strongly flowing water, which is also visible by an increase in grain size and sedimentary structures (e.g. cross-stratification). Other habitat categories used in the analysis were *low* and *high* energy (level 3), reflecting a broader categorization scheme, and the two contrasting habitat categories *lentic* and *lotic* (level 2). Further rock types representing the category *stagnant* are: coal, diatomite, dolomite, gyps lignite, marl, peal, phosphorite and tuff.

good preservation of the specimens, and differences in specimen completeness between habitats were calculated. Differences between taxonomic groups were also taken into account. Finding more complete specimens in a particular habitat would indicate a higher preservation probability and thus longer overall species durations. As the condition of specimens was not available for all data, respective samples of occurrences from each environment were used ($n = 81$, 72, 35, 14 for *stagnant*, *low-velocity*, *medium-velocity*, *high-velocity*, respectively). The specimen completeness between habitats was compared using Kruskal–Wallis rank sum test and Wilcoxon rank sum test for pairwise comparisons (fdr *p*-value correction).

2.3. Species duration, habitat and range size

Species durations were determined following Harnik [8]. We calculated the distance between geological stage mid-points of first and last occurrence of a species in millions of years, rounded to the next 1 million years. Species durations were compared between different taxonomic groups and habitats, as well as between habitats within each taxonomic group to control for phylogenetic dependence, using Kruskal–Wallis rank sum test and Wilcoxon rank sum test for pairwise comparisons (fdr *p*-value correction). To quantify the difference between those groups, we compared trimmed means. The trimmed mean is a robust estimate for the mean of an asymmetric distribution [35]. We used default settings provided by the describe function from the R psych package (trim = 0.1).

Fossil occurrences were assigned to different habitat categories based on their lithological context. As lithology reflects the sedimentary environment, and therefore the habitat in which the organism fossilized [36], we used lithology as a first order approximation of energetic regime and assigned each occurrence to one of four basic habitat categories. The assignment process is depicted in figure 1. We used three different sets (levels) of habitat categories: the first level comprised four distinct habitat categories, representing increasing energy in each depositional setting (*stagnant, low-velocity, medium-velocity, high-velocity*); a second level differentiated *lentic* from *lotic* depositional setting; and a third level indicated either a *low* or *high* energetic depositional setting. Level two and three assignments derived from level one data. We avoided redundant assignments and therefore bias caused by differing number of occurrences between species. When comparing species duration between habitats, the duration of a species got assigned once to the same habitat category on each level, regardless of the number of lithologies it occurred in. Therefore, a species with occurrences e.g. in claystone and shale was included only once in the category *stagnant*. The same principle was followed for the other habitat categories. Occurrences were excluded from analysis if they did not provide sufficient information to allow assignment to one of the four first level categories. This was the case for records with missing lithology data or data entries that did not allow inference on original energetic setting (e.g. 'cave infill'). Differences in durations between habitat groups were tested using Kruskal–Wallis rank sum test and Wilcoxon rank sum test for pairwise comparisons (fdr *p*-value correction).

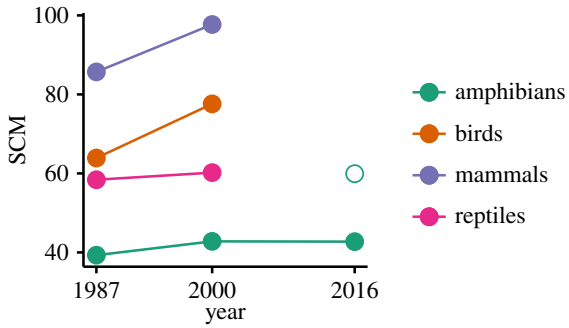

Figure 2. Comparison of simple completeness metric (SCM) for Cretaceous tetrapod groups from different years. Adapted from Fara & Benton [46] and completed with SCM values based on our dataset for 2016. SCM was calculated on the family level. The empty circle represents the total SCM value for amphibians; the solid ones give the value for Cretaceous families only.

We aimed at excluding phylogenetic dependence from purely environmental effects and therefore checked for differences between the durations and habitat preferences of the taxonomic groups. If species within some taxa naturally would have had longer/shorter durations than those of other taxa, and in addition would only occur in certain habitats, then longer durations in these habitats might not be caused by habitat, but simply be a result of one taxon dominating the habitat. A chi-squared test was used to test for differences between the expected and the observed frequencies of habitat categories within taxa. Differences in duration between habitat categories were compared within taxonomic groups.

Comparisons between taxonomic groups were always done on two taxonomic levels to account for the phylogenetic structure of the data, making sure to only compare taxa of the same hierarchical level. As Lepospondyli and Temnospondyli are both potential stem-group candidates for lissamphibians, we did comparisons between Lissamphibia groups (Allocaudata, Gymnophiona, Salientia and Urodela) and between the higher groups Temnospondyli, No-Temnospondyli (all species except Temnopondyli), Lepospondyli and No-Lepospondyli.

Geographical range size was included in our analysis as range size is a well-known influential factor for extinction risk in extinct and extant species [37,38]. Range size calculations followed the approach by Finnegan *et al*. [3] based on occupancy. A grid of 2×2 decimal degrees was projected on the palaeocoordinates of all occurrences and each occurrence assigned to a grid cell ID. The number of different grid cells, occupied by one species, resulted in the final grid cell count for each species. Geographical ranges were compared in the same way duration was compared between different habitat groups.

All analysis were done using the R environment v. 3.3.2 [39] with the additional packages ggplot2, readxl, gridExtra, psych, reshape and gsubfn [40–45]. An R script including analysis can be found in the electronic supplementary material.

3. Results

3.1. Quality and completeness

Our dataset comprises fossil amphibian occurrences from the Visean to the Holocene. Median geological stage duration was 5.7 million years with a median average deviation of 2.5 million years ($n = 53$).

The proportion of living amphibian families with a known fossil record was 33%, which is rather low compared to various groups of invertebrates and fishes (electronic supplementary material, figure S1). However, comparing amphibians with mammals showed that the preservation potential was smaller for amphibian families, but higher for amphibian genera and species (electronic supplementary material, table S3). The preservation probability of amphibians based on the duration frequency distribution showed that 33% of the amphibian species and about half of the genera were preserved at least once from one geological stage to the other (electronic supplementary material, table S3).

The SCM for species, genera and families showed values between 0.60 and 0.94 (electronic supplementary material, table S3). A comparison of Cretaceous amphibians with other vertebrate taxa placed their SCM at the lower end, yet slightly improving over time (figure 2, [46]). The SCM indicated that from all geological stages potentially containing amphibian fossils, at least 60% actually did contain fossil occurrences.

While testing for potential biases concerning the durations of species, we found no correlation of stage duration with species richness, number of single-interval species, or proportion of single-interval species ($|\rho| < 0.01$, $p > 0.95$, $n = 53$), indicating that geological stage duration had no biasing effect on the number of single-interval species. Ten per cent of PbDb species comprised occurrences tagged as coming from lagerstätten. Of those species, 69% were single-interval species, compared to 87% of single-interval species in the remaining, non-lagerstätten, dataset. Proportions of single-interval species thus differed significantly ($\chi^2_1 = 7.9$, $p < 0.01$), but were not biased towards lagerstätten. In addition there was no evidence for monographic effects, as publications describing a larger number of occurrences (more than 20) contained only 20% single-interval species in these occurrences. Lagerstätten and monographic effects were solely assessed for PbDb data, however we did not expect the two databases to differ in this regard as proportions of single-interval species occurrences were within comparative ranges in FosFARbase and PbDb (38% and 51%, respectively).

By comparing specimen completeness between different habitat categories we tested for habitat-dependent preservation potentials and for differences in this potential between taxonomic groups. For all amphibians, we found significant differences between the *stagnant* habitat category with both *low-velocity* and *medium-velocity* categories ($p < 0.05$), with a larger trimmed mean in *stagnant* (electronic supplementary material, table S4). The results were similar when we tested for differences individually within the groups Urodela, No-Lepospondyli, No-Temnospondyli and Temnospondyli. Specimen completeness also differed between the larger taxonomic groups, but not between lissamphibian groups (electronic supplementary material, table S5). Temnospondyli and Lepospondyli showed higher trimmed mean specimen completeness than other taxonomic groups. These results suggest a habitat influence, and, to some degree, a taxonomic influence on specimen completeness, with specimens from low-energy environments being more likely to be documented in the fossil record than specimens living in high-energy environments.

Based on these analyses on the quality and completeness of the fossil data, we are confident to use our data for the following analyses on differences in species' durations between habitats.

3.2. Habitat

Lithological information was available for 816 occurrences from 358 species. We investigated if durations of these species were connected to habitat type on three different levels (as defined in figure 1). On the first level, we found significant differences in species' durations among habitat categories (electronic supplementary material, table S6). Pairwise comparisons showed that species' durations in *medium-velocity* were longer than those from *stagnant* depositions ($p < 0.001$; figure 3; electronic supplementary material, table S6). Comparing durations from the broader categories *low/high* as well as *lentic/lotic* levels, showed *high* and *lotic* species to have longer durations than species from *low* or *lentic* settings ($p < 0.05$, figure 3, electronic supplementary material, table S6). These results indicate that species which lived in high-energy environments prevailed for longer periods than species from low-energy environments.

We tested for a phylogenetic signal in our data by comparing durations between taxonomic groups as well as between habitat preferences. We found significant differences between durations of Allocaudata and Salientia, with Allocaudata displaying longer median durations. Significant differences were also found between Temnospondyli and all other taxa (No-Temnospondyli), with No-Temnospondyli having a larger trimmed mean duration (figure 4*a*; electronic supplementary material, table S7).

For habitat preference, expected and observed frequencies of habitat categories did not significantly differ among lissamphibian groups, but among the taxonomic groups Lepospondyli, Temnospondyli, No-Lepospondyli and No-Temnospondyli ($\chi^2_9 = 30.565$, $p < 0.001$, figure 4*b*). Therefore, taxonomic identity and habitat usage were not entirely independent. Comparing the observed habitat frequencies with the expected frequencies revealed more than randomly expected occurrences of Lepospondyli in *stagnant* and less in *low-velocity* and *medium-velocity* habitats (electronic supplementary material, table S8).

As there were differences in the species duration, and to some extent in the habitat category frequencies between taxonomic groups, we checked for taxonomic influence in our observed duration pattern. We analysed the durations in habitat categories in all taxonomic groups individually. We detected significant duration differences among habitats for the groups Salientia, No-Temnospondyli and No-Lepospondyli on all environmental levels (except No-Lepospondyli in level 2, electronic supplementary material, table S9). Pairwise comparison for level 1 habitat categories revealed significant differences in the comparison groups *stagnant/low-velocity* and *stagnant/medium-velocity* for all three

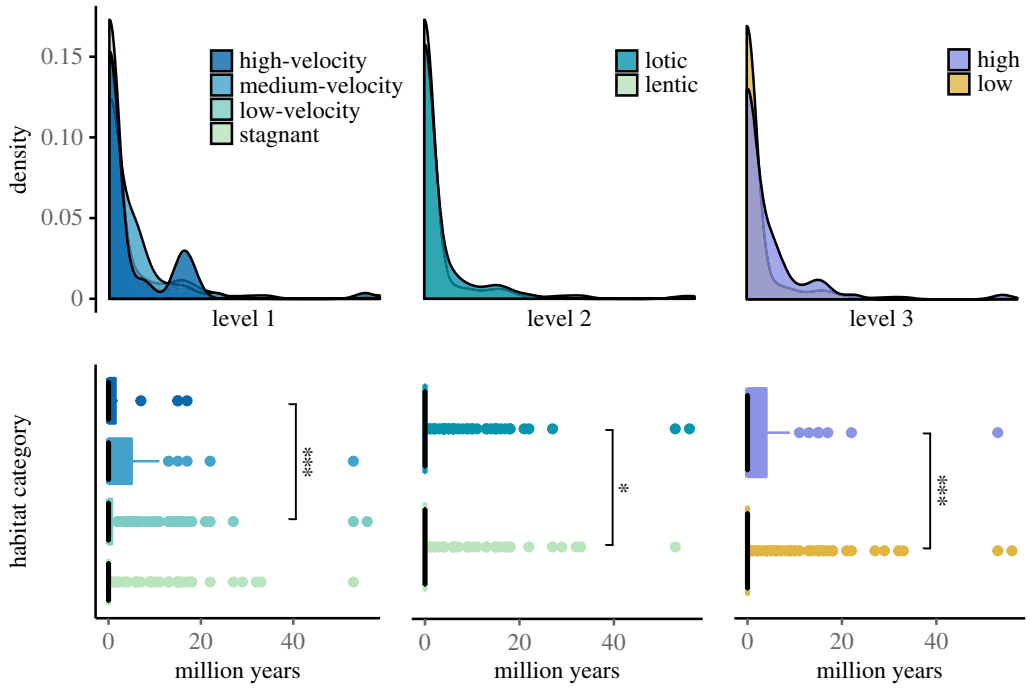

Figure 3. Durations of amphibian species in different environments. Species were grouped into four basic (level 1) and two broader environmental categories (level 2 and 3; compare figure 1). Sample sizes for groups were: *stagnant* (214), *low-velocity* (130), *medium-velocity* (56) and *high-velocity* (18); *lentic* (216) and *lotic* (176); *low* (319) and *high* (71). The upper panel shows the density distribution of durations (bandwidth = 2 million years), the lower panel shows the durations as boxplots, with black lines indicating the median and coloured areas illustrating the range between first and third quartiles. Significant differences are indicated by one, two and three asterisks indicating *p*-values smaller than 0.05, 0.01 and 0.001, respectively (for exact values compare electronic supplementary material, table S6). The largest outliers were caused by *Scapherpeton tectum* and *Gobiops desertus*, two extremely long living species.

taxonomic groups (except No-Lepospondyli *stagnant/low-velocity*, electronic supplementary material, table S10), with *low-velocity* and *medium-velocity* showing 1.1 to 3.5 million years larger trimmed mean durations compared to *stagnant* habitat. *Lotic* species had 1.1 to 1.9 million year longer trimmed mean durations than *lentic* species, and trimmed mean durations for *high-velocity* species were 1.3 to 1.8 million years longer than for *low-velocity* species. These results indicate that the pattern of longer durations in high-energy environments was strongest in Salientia, with Urodela and Temnospondyli showing a similar but non-significant trend.

While determining if geographical range might influence our results, we found significant differences between low and high-energy habitat categories on levels 1 and 3 ($p < 0.001$), with trimmed mean geographical ranges being larger in high-energy environments (electronic supplementary material, figure S2, table S11). The results showed that species from low energetic environments had a smaller geographical range size than species from high energetic environments. We observed similar results when controlling for geographical range size by analysing species with small and large range separately (electronic supplementary material, table S12).

Although we assume that most single-interval species in our data represent a real signal rather than being a result of preservation bias, we still wanted to consider the potential influence of false single-interval species. Therefore, we checked for equal distribution of single-interval species among habitats and tested for duration differences among habitats without single-interval taxa. Pearson's χ^2 test showed that proportions of single-interval species differed between habitat groups ($\chi^2_3 = 15.9$, $p < 0.01$), with *stagnant* showing more and *medium-velocity* showing less single-interval species than expected by chance (electronic supplementary material, figure S3, table S13). On level 3, habitat differences were significant as well ($\chi^2_1 = 16.2$, $p < 0.001$) with 87% of low energetic setting species being single-interval species compared to 66% in high energetic settings (electronic supplementary material, figure S3). Level 2 showed no significant habitat differences for single-interval species. We therefore concluded that single-interval species did overall occur more often in *low* energetic habitats. Excluding single-interval species from the duration pattern analysis among different habitats resulted in severe data reduction and the loss of the pattern observed in the complete dataset (electronic supplementary material, figure S4, table S14).

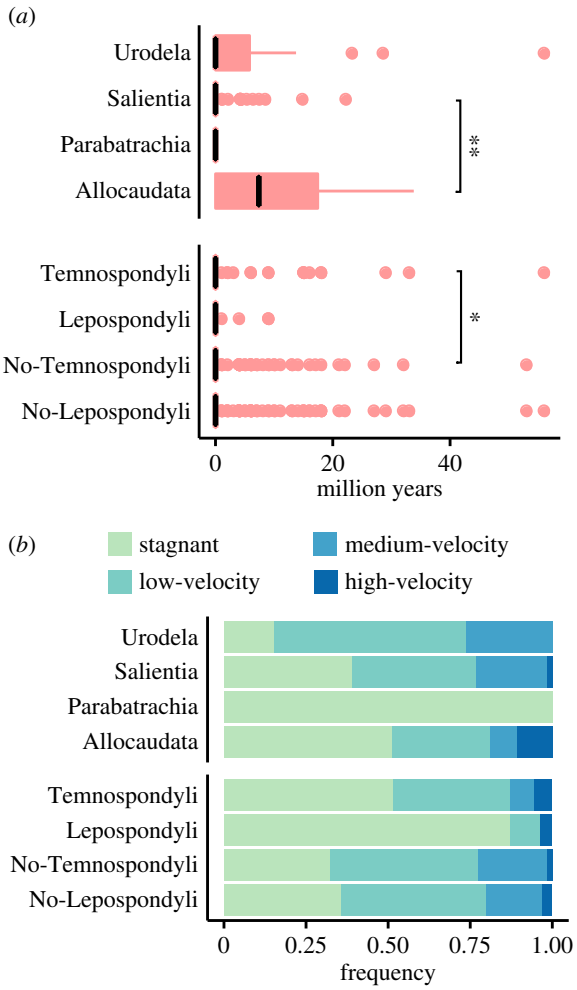

Figure 4. Durations and habitat preferences. (*a*) Durations of amphibian species from different taxonomic groups. Numbers of species for the groups were Allocaudata (11), Urodela (39), Parabatrachia (2), Salientia (80), No-Temnospondyli (171), No-Lepospondyli (310), Lepospondyli (39), and Temnospondyli (178). Black lines in boxplots indicate the median and coloured areas illustrate the range between first and third quartiles. Significance differences are depicted by one, two and three asterisks indicating *p*-values smaller than 0.05, 0.01 and 0.001, respectively. (*b*) Frequency of species in habitat categories for different amphibian groups, each species counted once per habitat category. Habitat categories as defined in figure 1. Parabatrachia had just two occurrences (from two species) which were from the same, *stagnant*, environment.

4. Discussion

Our analyses showed that extinct amphibian species differed in their duration depending on the habitat they lived in. Contrary to todays' situation, where extinction risk seems to be higher in species living and breeding in lotic waters [20], we found that extinct species persisted longer through time when they occurred in flowing water. An extensive quality analysis of our data not only revealed that amphibian data are at the lower end of available vertebrate fossil record quality (see also [30]), but also confirmed that the data were of sufficient reliability for our analyses. In particular these tests proofed that the most critical part of our data, the single-interval taxa, could be used as done herein.

4.1. Quality and completeness

Without knowing the weak spots of palaeontological data, the nonetheless fragmentary fossil record can easily lead to false inferences. Although we are well aware that the fossil record of amphibians is not perfect, we argue that for our approach the available data were sufficient.

To our knowledge the only studies hitherto explicitly examining the quality of the amphibian fossil record applied an SCM and showed a record with more gaps than mammals, birds and reptiles in the

Cretaceous period [30,46]. While our analysis confirmed these findings, we also found that the SCM for all time intervals was higher than in the Cretaceous (electronic supplementary material, table S3), which is also true for the complete tetrapod fossil record (fig. 2 in [30]). However, the SCM seems to be heavily influenced by the high number of single-interval taxa, naturally not containing any time gaps, and therefore potentially biasing the metric. Calculating the SCM on durations without their range endpoints gave much lower metric values (electronic supplementary material, table S3), which confirms the influence of single-interval species. The SCM was initially developed for usage on the family level [30]. However, we worked on the species level, and thus we considered testing the reliability of single-interval species to be more important than the SCM.

The preservation probability suggests the amphibian fossil record to be comparable to that of corals (electronic supplementary material, figure S1), which are often used for community structure analysis [47–49]. But comparing a terrestrial vertebrate taxon to a marine invertebrate taxon might not be very insightful owing to the usually much better preservation in marine sediments [50] and the vastly differing lifestyles of the taxa. The comparison of the amphibian data with those from mammals, being predominantly terrestrial vertebrates as well, instead showed the preservation potential of the amphibians to be lower on the family, but higher on the genus and the species levels (electronic supplementary material, table S3). This imbalance in preservation potential between taxon levels could hint at a family recognition problem, as morphological character differences between families are often subtle in amphibians and might have been lost owing to their mostly fragile nature. However, this does not affect our analysis as we solely focused on species durations.

More worrying was the high percentage of single-interval species in our fossil data, whose exclusion resulted in the loss of any significant difference in duration between habitats (electronic supplementary material, figure S4). Our analysis following Fitzgerald and Carlson [32] showed that the single-interval species are probably a real phenomenon and not produced by preservation or sampling biases. Dubey and Shine [51] support this view by giving a median age for extant anuran species of 1.5 million years, which is smaller than the length of most geological stages, and making it plausible for many species to be present in only one time interval. We also saw single-interval species being more common in low-energy habitats (electronic supplementary material, figure S3), which we assume to be a real signal, as we would expect the contrary given the better preservation potential in these environments. We argue that for our dataset the resulting loss in biodiversity and sample size would pose a greater bias to the results than occasional false single-interval species would do.

In accordance with our initial assumption of preservation potential being higher in calmer environments [36], we observed the highest specimen completeness in lentic habitats. This suggests that species from high-energy habitats should be affected the most by preservation bias and might have truncated durations as a result. As our high-energy records actually had longer durations, we conclude that the observed duration pattern is not because of a preservation bias. On the contrary, the effect of habitat on species duration might be even stronger than observed. On a taxonomic level, the higher overall specimen completeness detected for Temnospondyli and Lepospondyli is most likely attributed to the more robust morphology of Temnospondyli [52], and the preference of lentic water bodies by Lepospondyli (figure 4*b*). However, the higher preservation potential did not result in longer durations for Temnospondyli (figure 4). We initially assumed that good preservation results in potentially longer durations of species. In addition, differences in preservation potential of habitats might not only act on the duration but also on the total number of preserved species. An increase in preservation potential, like in low-energy habitats, might result in a larger proportion of short duration species, as fragile and rare species become more likely to be preserved in the first place. On the other hand, the preservation of all other species equally becomes more likely, and therefore their durations potentially longer. Depending on the ratio of newly added short duration species on one end and extended species durations on the other, a higher preservation potential might result in longer or shorter overall durations, or even no change at all. How many rare and fragile species become added to the preserved fauna with increasing preservation potential might depend on the morphological characteristics as well as abundance distribution of the species within the respective fauna. As clarification of this issue is beyond the scope of this study, we conservatively assume the effect of changing preservation potential between habitats to be neutral.

4.2. Habitat and species' duration

A basic assumption for our study is the correct assignment of occurrences to habitat categories. Out-of-habitat transportation might be a factor influencing our outcomes, as post-mortem transportation could

result in the species being assigned to the wrong habitat category. However, in a review on the quality of the fossil record Kidwell & Flessa [53] state that out-of-habitat transportation by secondary deposition is unlikely to occur. We have to admit, that we cannot judge what might be the more common scenario, dead animals washed from streams into ponds and lakes, or the converse way. Anyhow, based on the above review results and various biological data for many of our species, we are confident that our habitat categories reflect the actual habitats of the respective species.

More important for the interpretation of our main result, longer durations of species in high-energy habitats, indicating a lower extinction risk compared to their relatives in low-energy habitats, might be biases by phylogeny or other traits.

However, the phylogenetic influence on duration differences between habitats turned out to be rather small. Durations differed between few taxonomic groups (Temnospondyli and Allocaudata, figure 4*a*), which we expect to be of no influence to our main result as both groups did not differ in their habitat preference. The only group with a habitat preference that differed from the other taxa (Lepospondyli, figure 4*b*) showed again no difference in its duration pattern. Therefore, differences in duration between taxonomic groups did not co-occur with differences in the habitat preference of these groups. We also found the same duration differences between habitats in all except two taxonomic groups alone, which supports the phylogenetic independence of the results. In the two groups in which this trend was not statistically significant, Lepospondyli and Allocaudata (electronic supplementary material, table S9), we attribute the lack of any pattern to the low sample size in general and especially in high-energy environments for Lepospondyli.

A different trait influencing our results could be the geographical range size of species, which we found to have a positive correlation with the flow energy of species' habitats (electronic supplementary material, figure S2). As geographical range size is widely acknowledged as an important factor for extinction risk in amphibians and other taxa [38,54,55], this finding supports the lower extinction risk in high-energy habitats. However, controlling for geographical range size did not change the observed duration pattern between habitats, therefore it cannot be the only cause for our results. This result contrasts several findings from studies on insects with aquatic stages, showing that lotic species have smaller ranges than lentic species [56,57]. This is attributed to the lower temporal stability of stagnant water bodies and the resulting higher dispersal ability of inhabiting species [58]. One could argue that our observed larger ranges in high-energy habitats are the result of increased dispersal ability, caused by a widely spread water body with potentially active transportation. Another possible explanation for the reversed range size pattern between habitats might be the larger body sizes of early amphibians (Temnospondyli), which made species less prone to predation and thus able to inhabit larger, high-order streams, as well as possibly enabled them to simply move longer distances than smaller species.

In accordance with the larger range sizes in lentic species today (see above), it is usually assumed that amphibian species using low-energy habitats are less prone to extinction. Species breeding in ponds, for example, have to cope with lower habitat stability [58], which might make them more tolerant to environmental fluctuations. Higher extinction risk in lotic habitat species today [20] might be further caused by their mainly mountainous distribution, which naturally restricts their range size and isolates them from other areas with matching environmental conditions. It was also suggested that species associated with rivers might be more exposed to diseases [20].

To explain why a habitat type apparently changed from beneficial to detrimental for long species survival, one has to consider (i) changes in the habitat demands of amphibians over time, and (ii) changes in the habitat itself.

Habitat demands of a species are defined by various traits, for example morphology and life history. The most obvious change in morphology during amphibian evolution might be the overall decrease in body size. Comparing the 6 m length of the largest Temnospondyli with amphibians today [17], the latter are much smaller (the by far largest being *Andrias davidianus* with 1.5 m total length). However, current anuran species are not smaller than their ancestors, but still showed the inversed pattern in extinction risk between habitats. Therefore, the decrease in overall body size seems to be at least not the main reason for the change. The most prominent amphibian life-history trait, the biphasic life cycle with aquatic larvae, is assumed to be the ancient state for lissamphibians and supported by Temnospondyli fossils [23]. Moreover, the even more complex anuran metamorphosis with the apomorphic tadpole has been recorded since the mid Jurassic [59]. If we otherwise assume that a shift in life history from aquatic to more terrestrial lifestyles in the amphibian evolution might have caused the difference in duration pattern between past and present, we would expect to observe a difference between amphibian

orders, as these differ in their lifestyles, too. However, in our data, species across all orders displayed a higher extinction risk in lotic habitats.

If habitat demands of amphibians have remained basically unchanged, then changes in the habitat itself could be a reason for the observed differences in habitat-extinction risk relationship. It is thus tempting to assume an anthropogenic influence on that relationship, as anthropogenic effects on the environment clearly were not present in the fossil past. More specifically, there are several examples of anthropogenic factors influencing the amphibian fauna associated with rivers. Most importantly altered river structures and communities are an important factor that negatively affects amphibians. In various regions natural river systems, and in particular the abandoned channels and regularly flooded areas, are severely declining. On the other hand, exotic fish species have been released worldwide, potentially spreading diseases and increasing predatory pressure on tadpoles and adults [60]. Further logging activity along rivers has an indirect effect on the physical characteristics and the macroinvertebrate community of streams [61], which probably affects the amphibian community as well. More generally, disturbance of the habitat as measured by the amount of forested, agriculturally or residentially used area and concomitant alterations in water temperature, pH and dissolved oxygen have a negative influence on the relative abundance of stream-dwelling salamander populations [62]. These influences might act stronger on river habitats, which were and are of great importance to humans [63] and therefore strongly influenced [64]. Although we admit this to be speculative, we assume that most lotic species experienced more stable environmental conditions than their relatives living in lentic, often temporary habitats. The latter might be thus naturally already better adapted to frequent habitat changes. Lotic species in contrast might be particularly at risk by a multitude and increasing human-induced environmental changes, and the advantageous stability of this habitat over longer geological time scales has consequently been reversed.

5. Conclusion

We detected increased extinction risk in fossil amphibian species from low-energy water habitats, which objects today's situation. A trait character once favourable for a species turned into a disadvantage. Given that a likely reason is altered habitat conditions via anthropogenic influence, our work shows that the trait-environment interaction is an important factor to consider when learning about the influence of traits from the past. The fossil record might provide us with sort of a baseline, ancestral extinction risk, which obviously does not consider human influence. The differences between expected and observed influence on extinction risk might give us an insight about the underlying mechanisms of complex traits like habitat preference. When analysing the connection between traits and extinction risk, our results suggest to not only consider phylogenetic influences, but differences between temporal and environmental units too.

Ethics. This study required no ethical permit, since all data were retrieved from databases or the literature.

Data accessibility. R script and all data are available in the electronic supplementary material.

Authors' contributions. Both authors participated in the conception of the study and writing of the manuscript. M.T. collected the data and carried out the statistical analysis. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. M.T. was funded by Elsa Neumann Stipend (scholarship from the federal state of Berlin). M.-O.R. received no funding for this study.

Acknowledgements. We thank Florian Witzmann for advice on taxonomy and Martin Schobben for help with lithological questions and advice on figures. We thank the major contributors to the Paleobiology Database dataset John Alroy, Richard Butler and Matthew Carrano. We also thank Brandon Kilbourne for the language improvements and two anonymous reviewers for valuable comments and suggestions which have improved the quality of this work. This is Paleobiology Database publication number 281.

References

1. Dunham J, Peacock M, Tracy CR, Nielsen J. 1999 Assessing extinction risk: integrating genetic information. *Conserv. Ecol.* **3**, 2. (doi:10.5751/ES-00087-030102)
2. O'Grady JJ, Reed DH, Brook BW, Frankham R. 2004 What are the best correlates of predicted extinction risk? *Biol. Conserv.* **118**, 513–520. (doi:10.1016/j.biocon.2003.10.002)
3. Finnegan S *et al.* 2015 Paleontological baselines for evaluating extinction risk in the modern oceans. *Science* **348**, 567–670. (doi:10.1126/science.aaa6635)
4. Sodhi NS, Bickford D, Diesmos AC, Lee TM, Koh LP, Brook BW, Sekercioglu CH, Bradshaw CJA. 2008 Measuring the meltdown: drivers of global amphibian extinction and decline. *PLoS ONE* **3**, e1636. (doi:10.1371/journal.pone.0001636)
5. Van den Brink PJ, Alexander AC, Desrosiers M, Goedkoop W, Goethals PLM, Liess M, Dyer SD. 2011 Traits-based approaches in bioassessment and ecological risk assessment: strengths, weaknesses, opportunities and threats. *Integr. Environ. Assess. Manag.* **7**, 198–208. (doi:10.1002/ieam.109)

6. IUCN Standards and Petitions Subcommittee. 2014 *Guidelines for using the IUCN Red List categories and criteria*. Version 11. Prepared by the Standards and Petitions Subcommittee.
7. IUCN. 2012 *IUCN Red List categories and criteria: Version 3.1*, 2nd edn, iv + 32pp. Gland, Switzerland and Cambridge, UK: IUCN.
8. Harnik PG. 2011 Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proc. Natl Acad. Sci. USA* **108**, 13 594–13 599. (doi:10.1073/pnas.1100572108)
9. Powell MG. 2007 Geographic range and genus longevity of late Paleozoic brachiopods. *Paleobiology* **33**, 530–546. (doi:10.1666/07011.1)
10. Cardillo M, Huxtable JS, Bromham L. 2003 Geographic range size, life history and rates of diversification in Australian mammals. *J. Evol. Biol.* **16**, 282–288. (doi:10.1046/j.1420-9101.2003.00513.x)
11. Fritz SA, Bininda-Emonds ORP, Purvis A. 2009 Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecol. Lett.* **12**, 538–549. (doi:10.1111/j.1461-0248.2009.01307.x)
12. Jones KE, Purvis A, Gittleman JL. 2003 Biological correlates of extinction risk in bats. *Am. Nat.* **161**, 601–614. (doi:10.1086/368289)
13. Fröbisch NB, Olori JC, Schoch RR, Witzmann F. 2010 Amphibian development in the fossil record. *Semin. Cell Dev. Biol.* **21**, 424–431. (doi:10.1016/j.semcdb.2009.11.001)
14. Raup DM, Gould SJ, Schopf TJM, Simberloff DS. 1973 Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* **81**, 525–542. (doi:10.1086/627905)
15. Campbell KSW, Bell MW. 1977 A primitive amphibian from the Late Devonian of New South Wales. *Alcheringa An Australas. J. Palaeontol.* **1**, 369–381. (doi:10.1080/03115517708527771)
16. Young GC. 2006 Biostratigraphic and biogeographic context for tetrapod origins during the Devonian: Australian evidence. *Alcheringa An Australas. J. Palaeontol.* **30**, 409–428. (doi:10.1080/03115510609506875)
17. Vitt LJ, Caldwell JP. 2014 *Herpetology*. Oxford, UK: Elsevier Ltd.
18. Vié J-C, Hilton-Taylor C, Stuart SN (eds). 2009 *Wildlife in a changing world—an analysis of the 2008 IUCN Red List of threatened species*. Gland, Switzerland: IUCN.
19. Baillie JEM, Griffiths J, Turvey ST, Loh J, Collen B. 2010 *Evolution lost: status and trends of the World's vertebrates*. London, UK: Zoological Society of London.
20. Stuart SN, Hoffmann M, Chanson JS, Cox NA, Berridge RJ, Ramani P, Young BE. 2008 *Threatened amphibians of the World*. Barcelona, Spain: Lynx Edicions.
21. Alroy J, Butler RJ, Carrano MT. 2016 Taxonomic occurrences of Allocaudata, Amphibia, Anura, Caudata, Gymnophiona, Lepospondyli, and Temnospondyli recorded in the Paleobiology Database. Fossilworks. See http://fossilworks.org.
22. Böhme M, Ilg A. 2003 fosFARbase. Database of vertebrates: fossil fishes, amphibians, reptiles, birds. See www.wahre-staerke.com/.
23. Schoch RR. 2014 *Amphibian evolution: the life of early land vertebrates*. Chichester, UK: Wiley-Blackwell.
24. Frost DR. 2016 *Amphibian species of the World: an online reference. Version 6.0*. Electronic Database accessible at http://research.amnh.org/herpetology/amphibia/index.html. New York, USA: American Museum of Natural History.
25. Raup DM. 1972 Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071. (doi:10.1126/science.177.4054.1065)
26. Benton MJ, Dunhill AM, Lloyd GT, Marx FG. 2011 Assessing the quality of the fossil record: insights from vertebrates. *Geol. Soc. London, Spec. Publ.* **358**, 63–94. (doi:10.1144/SP358.6)
27. Schopf TJM. 1978 Fossilization potential of an intertidal fauna: Friday Harbor, Washington. *Paleobiology* **4**, 261–270. (doi:10.1017/S0094837300005996)
28. Foote M, Raup DM. 1996 Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* **22**, 121–140. (doi:10.1017/S0094837300016134)
29. Foote M, Sepkoski JJ. 1999 Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417. (doi:10.1038/18872)
30. Benton MJ. 1987 Mass extinctions among families of non-marine tetrapods: the data. *Mém. Soc. Géol. Fr. New Ser.* **150**, 21–32. (doi:10.1038/316811a0)
31. Benton MJ. 1999 Early origins of modern birds and mammals: molecules vs. morphology. *Bioessays* **21**, 1043–1051. (doi:10.1002/(SICI)1521-1878(199912)22:1<1043::AID-BIES8>3.0.CO;2-B)
32. Fitzgerald PC, Carlson SJ. 2006 Examining the latitudinal diversity gradient in Paleozoic terebratulide brachiopods: should singleton data be removed? *Paleobiology* **32**, 367–386. (doi:10.1666/05029.1)
33. Seilacher A, Reif W-E, Westphal F, Riding R, Clarkson ENK, Whittington HB. 1985 Sedimentological, ecological and temporal patterns of fossil lagerstatten [and discussion]. *Phil. Trans. R. Soc. Lond. B* **311**, 5–24. (doi:10.1098/rstb.1985.0134)
34. Benton MJ. 2008 How to find a dinosaur, and the role of synonymy in biodiversity studies. *Paleobiology* **34**, 516–533. (doi:10.1666/06077.1)
35. Marazzi A, Ruffieux C. 1999 The truncated mean of an asymmetric distribution. *Comput. Stat. Data Anal.* **32**, 79–100. (doi:10.1016/S0167-9473(99)00018-3)
36. Nichols G. 1999 *Sedimentology and stratigraphy*. Oxford, UK: Blackwell Science Ltd.
37. McKinney ML. 1997 Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annu. Rev. Ecol. Syst.* **28**, 495–516. (doi:10.1146/annurev.ecolsys.28.1.495)
38. Gaston KJ, Fuller RA. 2009 The sizes of species' geographic ranges. *J. Appl. Ecol.* **46**, 1–9. (doi:10.1111/j.1365-2664.2008.01596.x)
39. R Core Team. 2016 R: A language and environment for statistical computing. Vienna, Austria. See https://www.R-project.org/.
40. Wickham H. 2009 *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer Publishing Company, Incorporated.
41. Wickham H. 2016 Readxl: Read Excel Files. R package version 0.1.1. See https://CRAN.R-project.org/package=readxl.
42. Auguie B. 2016 GridExtra: Miscellaneous functions for 'Grid' graphics. R package version 2.2.1. See https://CRAN.R-project.org/package=gridExtra.
43. Revelle W. 2016 Psych: Procedures for psychological, psychometric, and personality research, Northwestern University, Evanston, Illinois, USA. See https://CRAN. R-project.org/package=psych Version=1.6.9.
44. Wickham H. 2007 Reshaping data with the reshape package. *J. Stat. Softw.* **21**, 1–20. (doi:10.18637/jss.v021.i12)
45. Grothendieck G. 2014 gsubfn: Utilities for strings and function arguments. R package version 0.6-6. See https://CRAN.R-project.org/package=gsubfn.
46. Fara E, Benton MJ. 2000 The fossil record of Cretaceous tetrapods. *Palaios* **15**, 161–165. (doi:10.1669/0883-1351(2000)015<0161:TFROCT>2.0.CO;2)
47. Knowlton N. 1992 Thresholds and multiple stable states in coral reef community dynamics. *Am. Zool.* **32**, 674–682. (doi:10.1093/icb/32.6.674)
48. Pandolfi JM. 1999 Response of Pleistocene coral reefs to environmental change over long temporal scales. *Am. Nat.* **39**, 113–130. (doi:10.1093/icb/39.1.113)
49. Bonelli JR, Brett CE, Miller AI, Bennington JB. 2006 Testing for faunal stability across a regional biotic transition: quantifying stasis and variation among recurring coral-rich biofacies in the Middle Devonian Appalachian Basin. *Paleobiology* **32**, 20–37. (doi:10.1666/05009.1)
50. Dunhill AM, Benton MJ, Twitchett RJ, Newell AJ. 2014 Testing the fossil record: sampling proxies and scaling in the British Triassic–Jurassic. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **404**, 1–11. (doi:10.1016/j.palaeo.2014.03.026)
51. Dubey S, Shine R. 2011 Geographic variation in the age of temperate-zone reptile and amphibian species: Southern Hemisphere species are older. *Biol. Lett.* **7**, 96–97. (doi:10.1098/rsbl.2010.0557)
52. Schoch RR, Milner AR. 2014 *Handbook of paleoherpetology—Part 3A2: Temnospondyli I*. München, Germany: Verlag Dr. Friedrich Pfeil.
53. Kidwell SM, Flessa KW. 1996 The quality of the fossil record: populations, species, and communities 1. *Annu. Rev. Earth Planet. Sci.* **24**, 433–464. (doi:10.1146/annurev.earth.24.1.433)
54. Howard SD, Bickford DP. 2014 Amphibians over the edge: silent extinction risk of data deficient species. *Divers. Distrib.* **20**, 837–846. (doi:10.1111/ddi.12218)
55. Kiessling W, Aberhan M. 2007 Geographical distribution and extinction risk: lessons from Triassic-Jurassic marine benthic organisms. *J. Biogeogr.* **34**, 1473–1489. (doi:10.1111/j.1365-2699.2007.01709.x)
56. Hof C, Brändle M, Brandl R. 2006 Lentic odonates have larger and more northern ranges than lotic species. *J. Biogeogr.* **33**, 63–70. (doi:10.1111/j.1365-2699.2005.01358.x)
57. Ribera I, Vogler AP. 2000 Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biol. J. Linn. Soc.* **71**, 33–52. (doi:10.1111/j.1095-8312.2000.tb01240.x)
58. Arribas P, Velasco J, Abellán P, Sánchez-Fernández D, Andújar C, Calosi P, Millán A, Ribera I, Bilton DT. 2012 Dispersal ability rather than ecological tolerance drives differences in range size between lentic and lotic water beetles (Coleoptera: Hydrophilidae). *J. Biogeogr.* **39**, 984–994. (doi:10.1111/j.1365-2699.2011.02641.x)
59. Yuan C, Zhang H, Li M, Ji X. 2003 Discovery of a middle Jurassic fossil tadpole from Daohugou region, Ningcheng, Inner Mongolia, China. *Acta Geol. Sin.* **78**, 145–148.

60. Gillespie G, Hero J-M. 1999 Potential impacts of introduced fish and fish translocations on Australian amphibians. In *Declines and disappearances of Australian frogs* (ed. A Campbell), pp. 131–144. Canberra, Australia: Biodiversity Group, Environment Australia.
61. Stone MK, Wallace JB. 1998 Long-term recovery of a mountain stream from clearcut logging: the effects of forest succession on benthic invertebrate community structure. *Freshw. Biol.* **39**, 151–169. (doi:10.1046/j.1365-2427.1998.00272.x)
62. Willson JD, Doreas ME. 2003 Effects of habitat disturbance on stream salamanders: implications for buffer zones and watershed management. *Conserv. Biol.* **17**, 763–771. (doi:10.1046/j.1523-1739.2003.02069.x)
63. Allan JD, Flecker AS. 1993 Biodiversity conservation in running waters. *Bioscience* **43**, 32–43. (doi:10.2307/1312104)
64. Everard M, Moggridge HL. 2012 Rediscovering the value of urban rivers. *Urban Ecosyst.* **15**, 293–314. (doi:10.1007/s11252-011-0174-7)

Paper 2

Evaluating the predicted extinction risk of living amphibian species with the fossil record

Tietje, M. and Rödel, M.-O. (2018), Evaluating the predicted extinction risk of living amphibian species with the fossil record. Ecology Letters 21(8): 1135-1142. doi:10.1111/ele.13080

ECOLOGY LETTERS

Ecology Letters, (2018) doi: 10.1111/ele.13080

LETTER

Evaluating the predicted extinction risk of living amphibian species with the fossil record

Melanie Tietje[1,2]* and **Mark-Oliver Rödel**[1,2]

[1]*Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science Invalidenstr. 43, 10115, Berlin, Germany*

[2]*Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB) Berlin, Germany*

**Correspondence: E-mail: melanie.tietje@mfn.berlin*

Abstract

Bridging the gap between the fossil record and conservation biology has recently become of great interest. The enormous number of documented extinctions across different taxa can provide insights into the extinction risk of living species. However, few studies have explored this connection. We used generalised boosted modelling to analyse the impact of several traits that are assumed to influence extinction risk on the stratigraphic duration of amphibian species in the fossil record. We used this fossil-calibrated model to predict the extinction risk for living species. We observed a high consensus between our predicted species durations and the current IUCN Red List status of living amphibian species. We also found that today's Data Deficient species are mainly predicted to experience short durations, hinting at their likely high threat status. Our study suggests that the fossil record can be a suitable tool for the evaluation of current taxa-specific Red Listing status.

Keywords

Amphibians, conservation, Data Deficient, extinction risk, fossil record, prediction, traits.

Ecology Letters (2018)

INTRODUCTION

The currently living amphibians comprise a very high proportion of endangered species, due to factors like habitat degradation and loss, pollution and disease (Rouse *et al.* 1999; Daszak *et al.* 2003; Cushman 2006). The risk of extinction due to these factors is, to some degree, shaped by their ecological and morphological traits. The influence of traits like body size, clutch and range size or several environmental variables, like variations in precipitation and temperature, has been analysed in several studies (Cooper *et al.* 2008; Sodhi *et al.* 2008; Hirschfeld & Rödel 2017), and partly used to predict species' extinction risk (Howard & Bickford 2014).

These studies of extinction risk usually combine trait data with the present-day extinction risk of a species as it has been assessed in the IUCN Red Lists (Purvis *et al.* 2000; Sodhi *et al.* 2008; Howard & Bickford 2014). These assessments are based on quantitative data as well as on expert opinions. The most important criteria are range size and population trend, the latter most often being based on assumptions rather than on data (Sodhi *et al.* 2008; IUCN 2012). The influence of these factors on species' current extinction risk is lastly derived from logical arguments and to a lesser degree on real observation; in particular, as we usually lack the ability to track extinction events because they operate over longer time scales than studies cover, or occur in remote areas where they remain undetected (IUCN 2017). The general validity of the IUCN criteria thus seems logical, but remains uncertain. As recommended already by Mace *et al.* (1992), especially the quantitative Red List criteria should be verified using data on species' extinctions. The more we know about the connection between certain characteristics of a species and its extinction risk, the more reliable the assessment results become, which would finally lead to better resource allocation and more effective conservation actions.

Herein we hypothesise, in accordance with, for example, a review by Dietl & Flessa (2011) and an opinion paper by Fritz *et al.* (2013), that the fossil record can be used (1) to provide (an additional) form of verification for the existing assumptions about the correlation of particular traits with the extinction risk of species, and (2) to identify traits that influence the extinction risk of species, but are not yet incorporated in risk assessments. The combination of conservation and palaeontology has the potential to strengthen the reliability of the Red List assessments and might also in particular offer an urgently needed tool in predicting extinction risk of unassessed or Data Deficient species (Bland *et al.* 2014; Howard & Bickford 2014; Jetz & Freckleton 2015). Such combination of fossil record and neontological data has rarely been practised, with those studies being conducted focusing on (mostly invertebrate) marine taxa or the influence of single traits on extinction risk (Finnegan *et al.* 2015; Tietje & Rödel 2017a).

Here, we use a fossil-based model to predict the extinction risk of living amphibian species. We show that the extinction risk, predicted by our model, matches the IUCN Red List status well for most species. We also predict that species with insufficient data for assessment (Data Deficient) have an on average even higher extinction risk than species classified as Vulnerable.

MATERIALS AND METHODS

Data

We collected a dataset containing six different variables: species stratigraphic range (duration), abundance, geographic range size, latitude of occurrences and body size. Taxonomic groups as used in this study were Allocaudata (identical with Albanerpetontidae Fox & Naylor, 1982), Lepospondyli, Parabatrachia (= Gymnophiona and their stem taxa), Salientia

(= Anura and their stem taxa), Temnospondyli and Urodela (= Caudata and their stem taxa). Amphibia, as herein understood, therefore comprise all Lissamphibia including their stem taxa as well as Temnospondyli and Lepospondyli to account for the temnospondyl- and lepospondyl hypothesis respectively (see Tietje & Rödel 2017a for more detailed information). Data and R code to replicate this study are available on https://doi.org/10.5281/zenodo.1206123.

Extinct species

The data for 354 extinct species were collected from the literature and databases of fossil occurrences, with no restrictions on time or area. The data were downloaded from the Paleobiology Database (PBDB) on 19 August 2014 via http://fossilworks.org (Alroy *et al*. 2016) and updated with a newer version on 15 February 2016. Abundance and body size data for the included species were added from literature and the fosFARbase (http://wahre-staerke.com; Böhme & Ilg 2003). The data processing is illustrated in Fig. 1. Values for species duration were measured as distance between the midpoints of the stratigraphic unit in which a species occurred first and last, and then rounded to the next full million years. The stratigraphic units of our data are chronostratigraphic stages as provided by the PBDB, the median stage length was 5.3 million years (n = 51, Fig. S1).

Geographic range measures were calculated for each geologic stage and species, the final value for the species is the maximum value of all geologic stages. Geographic range size and latitudinal range were calculated as the maximum great circle distance and maximum latitudinal distance between species occurrences, respectively. Abundance was based on the minimum number of individuals and specimen counts per locality and geologic stage

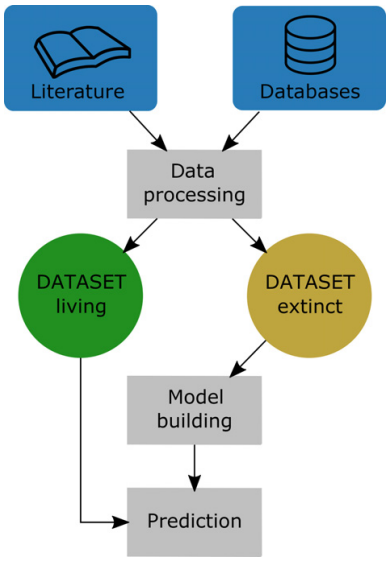

Figure 1 Data processing and modelling chart. Data processing includes the assignment of abundance categories, the calculation of geographic measurements and body sizes.

(specimens are here single fossils, i.e. an articulated skeleton, or a single bone). The maximum values per species were clustered (k-means clustering) into four numeric, ordinal categories (1, 2, 3 and 4). Body size was collected as maximum snout-vent-length (SVL) and, if not available, derived from total length and absolute skull length via linear regression models (see Table S1 in supporting information). All variables are summarised and further explained in Table S1. Basic statistics summarising the number and latitudinal distribution of occurrences, as well as a map of occurrences, can be found in the supporting information (section data summary).

Of the 354 fossil species, 165 species had complete datasets. The other species were lacking either abundance or body size data, or both. We decided to complement these missing data to get more substance to our model, and to avoid losing information from incomplete species. We therefore performed data imputation for missing data using multivariate imputation by chained equations from the mice R package (Van Buuren & Groothuis-Oudshoorn 2011), which creates multiple imputed datasets depending on different sampling sets of the other available variables, and takes the mean of these datasets as finally imputed value. To assess the quality of the imputed values, we ran diagnostic checks on the imputed data following Van Buuren & Groothuis-Oudshoorn (2011), which includes checking for convergence of the imputing algorithms, plausibility of imputed data, and occurrences of impossible data. For further method details and discussion on the possible influence of data imputation, see section 'Data imputation' in supporting information.

Living species

The data for 1382 living species were collected using the IUCN Red List, AmphibiaWeb and the literature. The number of species here was limited by the availability of body size data, which were collected as snout-vent-length from Hirschfeld & Rödel (2017), Trochet *et al*. (2014), Ruland & Jeschke (2016) and AmphibiaWeb (2017). The dataset resembles the taxonomic proportions of extant anurans and caudates; however, there were no caecilians included as data for this taxon was scarce, that is, most species assessed as Data Deficient.

The current, assessed extinction risk was extracted from the IUCN Red List on 3 February 2017. Abundance was extracted by text data mining from the population descriptions of the species in the IUCN Red List on 3 February 2017, using several keywords like, for example, 'common' or 'rare' (see Table S3 for a comprehensive list of keywords). According to their keywords species were assigned to one of four abundance categories, with one being the category for the rarest species. Geographic range measures were extracted from the shapefiles provided by the IUCN for each species (downloaded on 11 July 2016), using the rgdal and fields R packages (Nychka *et al*. 2015; Roger *et al*. 2017).

METHODS

Model building

Data for the extinct species were used to fit models correlating stratigraphic duration to the following traits: abundance,

geographic range size, latitudinal position and body size. Habitat type was excluded from analysis, as previous work had shown that the influence of the habitat type is not stable over time (Tietje & Rödel 2017a).

We built three different types of models: A generalised additive model (GAM), a randomForest (rF) and a generalised boosted model (GBM). The model type which performed best in cross-validation was fitted again to several subsets of the data, including a model fitted to the dataset excluding single-interval species to account for possible effects of our imbalanced response variable, and a randomised version of the data to obtain a null model. The best model trained on the extinct species data was then applied on the dataset of living species to predict duration estimates for living species in million years.

We used the caret R package for model training and comparison (from Kuhn *et al*. 2017). The model parameters were optimised using three separate 10-fold cross-validation on each possible set of parameters, meaning the extinct species dataset was randomly split into 10 equal-sized subsamples, from which nine subsamples were used as training data and one was retained as validation data for testing the model. This process was repeated until each of the subsamples was used once as validation set. The procedure was repeated three times. The mean RMSE (root mean squared error) from cross-validation resamples was compared for each set of parameters, choosing the parameter set with the minimum mean RMSE as the final model (Fig. S7). Model performance was estimated with RMSE and R^2, which were obtained from cross-validation of the final model and used for model comparison.

We observed slight differences in the mean stratigraphic ranges between the extinct amphibian taxa (Fig. S13), which might cause performance differences for the model between taxa. For example, while anuran durations might be overpredicted by the model, the opposite could occur for caudates. To reduce this potential taxonomic bias, we followed the approach by Finnegan *et al*. (2015) and built an additional GBM which was adjusted to these taxonomic differences. We cross-validated our GBM performance via bootstrap, using half of the data for model building and half of it for prediction. We then fitted a model on the predicted vs. observed durations, which was used to correct the initially predicted durations for potential taxonomic bias.

Duration predictions for living species

To compare the prediction results with the current species IUCN assessment results, the predicted duration values were grouped according to their species' corresponding IUCN Red List category (for category definition see IUCN 2012). These groups were checked for significant differences in their mean duration using Kruskal–Wallis rank sum test and Pairwise Wilcoxon rank sum test.

RESULTS

Model evaluation

We found the GBM to be the most suitable model for relating the duration of extinct species to their traits, which explained on average 26% of the variability of the duration of extinct species (Fig. S19 and Table S5). As assessed via cross-validation, the GBM outperformed the rF and GAM with a mean RMSE of 5.14 and a R^2 of 0.26 (based on the three times 10-fold cross-validation). The best model parameters for the

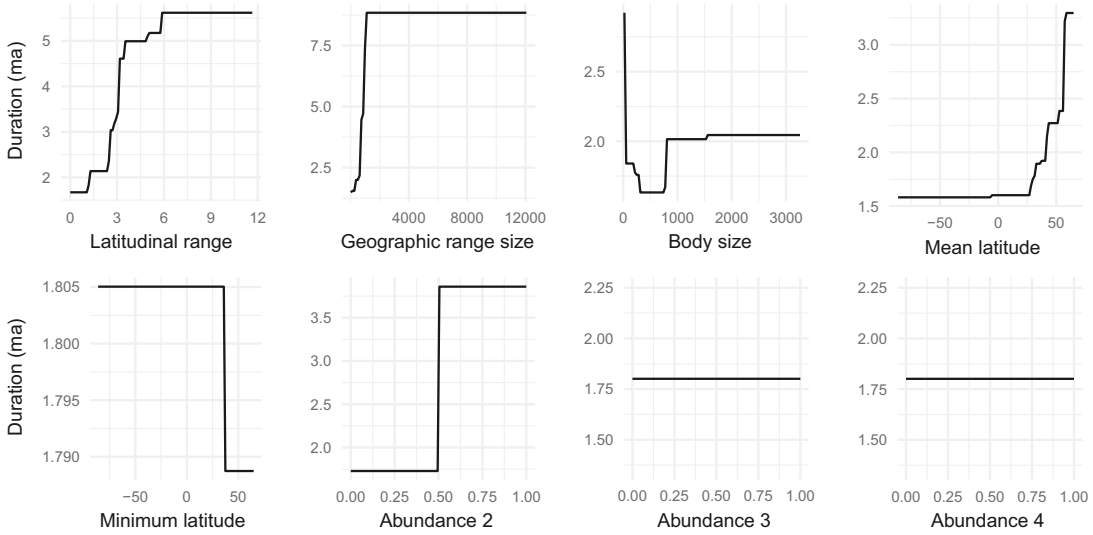

Figure 2 Marginal effects of each variable on species' durations in the generalised boosted model (GBM). Holding all other variables constant, each plot shows the average predicted duration for each predictor variable value in the model.

GBM were 200 trees, an interaction depth of 1, a shrinkage of 0.01 and a minimum number of observations in the trees terminal nodes of 5 (supporting information Output 5 and Fig. S10). We also found that the GBM fitted with the complete extinct species dataset yielded better results in cross-validation than GBMs fitted on subsets of the data (Fig. S19 and Table S5). All subset models performed worse, except for the Lissamphibia model. This model performed better when assessed with RMSE and R^2, but had a lower sample size and higher standard deviations in the model metrics assessed in cross-validation, which was also the case for the GBM fitted on the dataset excluding all incomplete occurrences. The prediction quality was more variable in these two models, which is why we chose the complete GBM as final model.

The variable importance of the final GBM (Table 1) showed that the geographic range size has the strongest influence on the duration of a species, followed by latitudinal range, mean latitude and body size. Minimum latitude and abundance were of least importance concerning duration in the GBM (see Fig. 2).

We detected performance differences of the GBM between different taxonomic groups (Fig. S14). Cross-validation of the model showed that durations for Salientia, Temnospondyli and Lepospondyli tended to be overestimated by the model, whereas durations for Allocaudata and Urodela species were overestimated for short durations as well, but underestimated for longer durations. This result resembles the generally longer mean durations of Allocaudata and Urodela in comparison with the other taxonomic groups. However, when we adjusted the predictions for living species for these taxonomic differences, the correction did not change the basic prediction pattern (Fig. S20).

The null model explained on average only 3% of the variability in the extinct species duration (Table S5). Repeated fitting ($n = 50$) of a GBM for different randomizations yielded similar results (Fig. S18). The null model therefore only explained very few of the variation in the duration of extinct species, with its R^2 being significantly smaller than all other models (t-test, $P < 0.001$, Output 8).

Table 1 Variable importance for the generalised boosted model (GBM), showing the relative influence of the predictor variables on the duration of amphibian species

Variable	Relative influence
Geographic range size	54.35
Latitudinal range	16.43
Mean latitude	11.52
Abundance 2	8.97
Body size	8.74
Minimum latitude	0.00
Abundance 3	0.00
Abundance 4	0.00

The relative influence is an important measure for the influence of each predictor variable in the model, calculated as described in Friedman (2001). Values are scaled to match 100%. The geographic range size (calculated as great circle distance) had the strongest influence on the model, followed by latitudinal range and mean latitude. For abundance category descriptions, see Table S1.

Prediction results

Our fossil-calibrated model was then used to predict the durations of living amphibian species. Our model predicts that duration for species decreases successively from Least Concern (LC) to Extinct (EX), with LC species showing a median predicted duration that is more than two times the median duration of Near Threatened (NT) species and more than four times the median duration of all other IUCN Red List categories (Fig. 3). The duration of the LC species is predicted as significantly higher than all other groups ($P < 0.05$, Output 6 in supporting information). Data Deficient species (DD) showed a median predicted duration that was most similar to the predicted durations of Critically Endangered (CR) species.

Predicting the durations of living species using any of the other model types or subsets of the data (excluding the null model) results in similar duration patterns (Fig. S20).

DISCUSSION

IUCN Red List assessments are fundamental for conservation research and planning (Hoffmann *et al.* 2008) and aim at providing an objective framework for the classification of a broad range of species (IUCN 2012). Testing the framework of the classification criteria is of high importance though (Mace *et al.* 1992).

Based on traits of fossil amphibian species, we predicted durations of living species and compared them to the respective IUCN Red List status. Predicted durations indeed well reflected these categories, decreasing continuously from Least Concern to higher extinction risk categories. Data Deficient species ranked among those being rated as Critically Endangered.

We have chosen to use the GBM as our final model as it allows for nonlinear relationships between predictor and response variable and is robust towards interaction effects between variables (Elith *et al.* 2008). Our model explained 26% of variation in the duration of fossil species, which seems reasonable given the unavoidable inaccuracies of fossil data and the complex interplay of factors that can influence the extinction risk for species (Newell 1959; Sepkoski *et al.* 1981; Purvis *et al.* 2000; Kidwell & Holland 2002; Harnik 2011).

Geographic range size was the most influential factor among the tested variables in our GBM. These results are supported by Sodhi *et al.* (2008), Cooper *et al.* (2008) and Hirschfeld & Rödel (2017), who found geographic range size to be of most importance for the extinction risk of extant amphibian species. The importance of geographic range measures is also accounted for in the rating procedures and threat categories within the IUCN assessment process of species (IUCN 2012). We observed with our model that the importance of geographic range measurements was not restricted to extant species, but most likely is a consistent factor for extinction risk. Body size and abundance were of minor importance in our model, which is matching other studies on the fossil record and contemporary amphibian species (Sodhi *et al.* 2008; Harnik 2011; Harnik *et al.* 2012). This observed pattern in variable importance was also supported by the rF and GAM (abundance was not assessed in the GAM due to the factorial nature of the variable though) and the GBMs on the

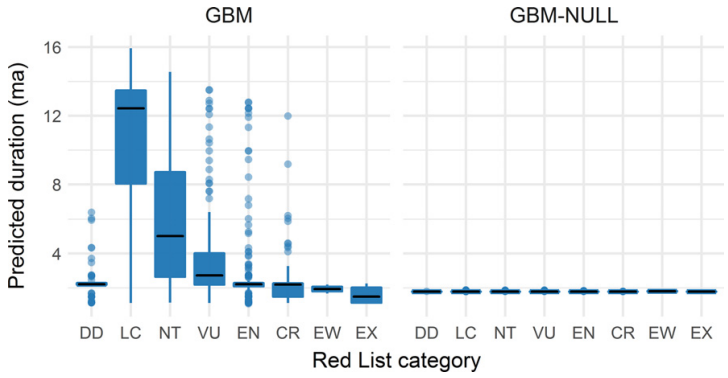

Figure 3 Predicted durations in million years for living amphibian species, based on the GBM models fitted to extinct species. Left: final GBM, right: Null model. Model parameters were *n* trees = 200, interaction depth = 1, shrinkage = 0.01 and minimum number of observations in the trees terminal nodes = 5. Predicted durations are grouped by their species' IUCN Red List assessment status. GBM = Generalised boosted model, GBM-NULL = Null model on randomised data. Red List categories (with number of species) are Data Deficient (66), Least Concern (740), Near Threatened (98), Vulnerable (151), Endangered (192), Critically Endangered (128), Extinct in the Wild (2), Extinct (5).

data subsets Lissamphibia and single-interval species (Figs S15 and S16).

In a previous study (Tietje & Rödel 2017a), we showed that the influence of traits on species' duration can change over time, specifically the habitat type, which was the reason why we excluded habitat type from our model. Reasons for this change might be a combination of the strong human impact on the environment in combination with an assumed minor ability of lotic amphibians to deal with rather quick changes of the environment. However, we assume the influence of geographic range measures to be stable over time. Studies on the influence of range size on a species survival suggest that its influence strength is weaker during mass extinction events compared to background extinction; however, the influence never becomes negative (Jablonski 2005; Payne & Finnegan 2007). Body size, however, might have a more flexible influence on the duration of species, as suggested by partly contrasting study results on extant amphibian species which showed that a larger body size was either bad or neutral for a species extinction risk, but never positive (Williams & Hero 1998; Lips *et al.* 2003; Sodhi *et al.* 2008; Hirschfeld & Rödel 2017). A possible scenario to inverse this effect direction could be a substantially warmer climate, in which large bodies might prevent desiccation better than smaller ones. However, extant amphibians in the tropics tend to be smaller than their relatives from colder areas (Lindsey 1966). As the influence of body size is rather small in our model, we do not expect potential fluctuations in the effect direction of body size to change our prediction patterns.

The predicted durations for the different Red List categories appeared well separated and let us conclude that our fossil species model gives a reasonably good estimation for the extinction risk in living species. Given the different nature of the applied models and the data subsets, we are confident to assume our predictions are the results of the interaction by the used prediction factors with duration. In addition, the null model revealed how results look like in case of no correlation between the variables and duration (Fig. 3). The correction for the taxonomic difference in our model (Fig. S13) mostly shortened the predicted durations in general, but did not change the basic pattern (Fig. S20). A potential phylogenetic signal could superpose the influence of the traits in our model, making duration the result of shared evolutionary history rather than the result of the trait combinations. We cannot rule out that this effect might weaken the association of the used traits and duration; however, several studies show that accounting for phylogeny, especially regarding geographic range size (our most important trait), only moderately weakened the association (Crampton *et al.* 2010; Hopkins & Klingenberg 2011; Harnik *et al.* 2014). Overall our model predictions well mirrored the IUCN red listing procedure. Most notably, we predicted the Data Deficient species to be at high risk, which makes intuitively sense given their presumed small ranges and low population numbers. This is also supported by Howard & Bickford (2014) and Morais *et al.* (2013), who predicted that 63 and 57%, respectively, of DD species are threatened.

However, the combination of fossil and modern data bears some potential pitfalls (Fritz *et al.* 2013). We would assume that the data gathered from fossil species usually depicts underestimates of their real values due to preservation restrictions of the fossil record (Darwin 1859; Newell 1959; Kidwell & Holland 2002). On the other hand fossil data represents time-averaged measures, like abundance or geographic range size, as fossils from one geologic stage might stem from organisms living at slightly different times, which are not resolved in the geological record (Kowalewski 1996). This might lead to an overestimation of, for instance, geographic range size in cases where, for example, the area occupied by a species shifted through time, but due to time-averaging the range appeared larger. A crucial point for our approach was the duration estimation for species, which serves as the fossil record's equivalent for extinction risk in our analysis. As neither the first nor the last occurrence of a species will be

recorded in the fossil record (Signor-Lipps effect), estimated durations will be consequently too short (Signor & Lipps 1982). However, we assume this effect to equally impact our data as we did not find evidence for differences in preservation between habitats or differently sized species Tietje & Rödel (2017a). Our model gave a weaker but still comparable prediction pattern when fitted on a data subset excluding all single-interval species (Fig. S20), as it is common practice in palaeontology to exclude duration biases. The skewed frequency structure of our response variable (Fig. S13b) does not lead to a biased prediction pattern that favours short-duration species. Therefore, we saw no reason to exclude single-interval species from our analysis.

The most common objections concerning the integration of fossil and contemporary data are doubts about (1) the comparability of effects of stressors acting in the past and today, and (2) the differing timescales of contemporary and fossil data. While some current drivers of extinction risk, like climatic changes, have their analogues in the fossil record, others like overexploitation and pollution are regarded to come without a deep-time analogue in the fossil record (Harnik *et al.* 2012). However, overexploitation due to pet trade and consumption affects only few species on the large scale (3 and 2%, respectively, Ruland & Jeschke 2016; Altherr *et al.* 2011), and might be mirrored in the fossil record by the arrival of new predators or diseases. Pollution is a stressor that in its human-made form is unique; however, we assume pollution events to have occurred in the past as well, caused by, that is, volcanic activity. Therefore, we argue that for todays' human-induced threats there are actually potential natural counterparts, and that stressors are therefore comparable.

The resulting extinction risk, as represented by the IUCN Red List, is a snapshot in the history of a species and depicts just its latest development. Our model, however, originates from a much broader timescale, with a resolution of several million years. A potential problem of this multiscale approach could be the inability of our model to consider temporal fluctuations in some species traits, such as, for example, geographic range size or abundances. A (current) temporarily small geographic range size would be assumed to be the maximum distribution, which would automatically lead to an underestimate of predicted durations. Dramatic shifts in range size are known from the fossil record, for example, formerly rare and range-restricted species becoming abundant and widespread after a mass extinction event. As most amphibian species seem to have very particular habitat requirements, we assume that such dramatic range shifts are however rare and negligible. However, even when keeping these potential problems in mind, the congruence of our fossil-based model with the contemporary assessments of extinction risk seems to show that the included factors do indeed act comparable on different timescales.

One potential shortcoming of our analysis might be the different nature of extinction risk measures in palaeontological and neontological data. As we use duration as a measure of extinction risk in the fossil dataset, the GBM cannot be used to predict a Red List category for living species, but delivers a duration estimate. However, comparing these estimates between species from the same Red List category and detecting large discrepancies in predicted durations, may even hint at possible misclassifications in the Red List, given our prediction works correctly. Keeping in mind the monetary and manpower restrictions that apply to both assessment and conservation efforts, and the rapid changes in our environment, it seems advisable to concentrate assessment and subsequently conservational efforts on species where both a comparatively low predicted duration and an endangered category or data deficiency concur. This applied to 21% of species included in the Red List category EN, which were predicted with shorter than average durations (Table S6). However, 8–16% of species showed longer than average predicted durations in the categories VU, EN and EX. Regarding the broad range of our predictions, especially for LC and NT species, it might be advisable to have a closer look at the shortest duration species in these groups as well.

Our model supports the importance of geographic range size for a species' extinction risk. Furthermore, the influence of range size seems to be stable over long time periods, as this model emphasises. Based on amphibians from different time periods and taxonomic groups, we could confirm that the most common Red List assessment procedures perform well on modern species. In addition, we conclude that species from the fossil record can indeed add to our knowledge about extinction risk-influencing traits of living species by providing various cases of extinction events. The very short durations predicted for the 66 Data Deficient species reflect the high risk these species are likely facing and emphasise the need for detailed assessment and, if proven needful, conservation actions.

ACKNOWLEDGEMENT

We thank the numerous authors of the original studies who provided the source data on which this study is based, and the many data enterers of PBDB, particularly the main contributors John Alroy, Richard Butler and Matthew Carrano. This is Paleobiology Database publication number 309. MT was supported in this study by the Elsa-Neumann Fellowship of the Federal State of Berlin and the fellow programme 'Freies Wissen' by Wikimedia Germany and the Stifterverband. Gregory Dietl and an anonymous reviewer provided very useful critiques, improving the manuscript. Florian Witzmann provided help with the collection of body size data.

AUTHORSHIP

Data collection, modelling and analysis were performed by MT. Both authors contributed substantially to the design of the study and writing of the manuscript.

DATA ACCESSIBILITY STATEMENT

All data and R code to replicate this study are available on https://doi.org/10.5281/zenodo.1206123 (Tietje & Rödel 2017b).

REFERENCES

Alroy, J., Butler, R.J. & Carrano, M.T. (2016). Taxonomic occurrences of Allocaudata, Amphibia, Anura, Caudata, Gymnophiona, Lepospondyli, and Temnospondyli recorded in the Paleobiology Database. *Fossilworks*. Available at: http://fossilworks.org. Last accessed 15 February 2016.

Altherr, S., Goyenechea, A. & Schubert, D. (2011). Canapés to extinction. A report by Pro Wildlife, Defenders of Wildlife and Animal Welfare Institute, 1–33.

AmphibiaWeb (2017). University of California, Berkeley, CA, USA. Available at: https://amphibiaweb.org/. Last accessed 10 February 2017.

Bland, L.M., Collen, B., Orme, C.D.L. & Bielby, J. (2014). Predicting the conservation status of data-deficient species. *Conserv. Biol.*, 29, 250–259.

Böhme, M. & Ilg, A. (2003). fosFARbase. Database of Vertebrates: fossil Fishes, Amphibians, Reptiles, Birds. Available at: www.wahre-staerke.com/. Last accessed 18 July 2016.

Cooper, N., Bielby, J., Thomas, G.H. & Purvis, A. (2008). Macroecology and extinction risk correlates of frogs. *Glob. Ecol. Biogeogr.*, 17, 211–221.

Crampton, J.S., Cooper, R.A., Beu, A.G., Foote, M. & Marshall, B.A. (2010). Biotic influences on species duration: interactions between traits in marine molluscs. *Paleobiology*, 36, 204–223.

Cushman, S.A. (2006). Effects of habitat loss and fragmentation on amphibians: a review and prospectus. *Biol. Conserv.*, 128, 231–240.

Darwin, C.R. (1859). On the imperfection of the geological record. In: *The Origin of Species*. John Murray, London, UK, pp. 279–311.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2003). Infectious disease and amphibian population declines. *Divers. Distrib.*, 9, 141–150.

Dietl, G.P. & Flessa, K.W. (2011). Conservation paleobiology: putting the dead to work. *Trends Ecol. Evol.*, 26, 30–37.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.*, 77, 802–813.

Finnegan, S., Anderson, S.C., Harnik, P.G., Simpson, C., Tittensor, D.P., Byrnes, J.E. *et al.* (2015). Paleontological baselines for evaluating extinction risk in the modern oceans. *Science*, 348, 567–670.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.*, 29, 1189–1232.

Fritz, S.A., Schnitzler, J., Eronen, J.T., Hof, C., Böhning-Gaese, K. & Graham, C.H. (2013). Diversity in time and space: wanted dead and alive. *Trends Ecol. Evol.*, 28, 509–516.

Fox, R.C. & Naylor, B.G. (1982). A reconsideration of the relationships of the fossil amphibian Albanerpeton. *Can. J. Earth Sci.*, 19, 118–128.

Harnik, P.G. (2011). Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proc. Natl Acad. Sci. U. S. A.*, 108, 13594–13599.

Harnik, P.G., Simpson, C. & Payne, J.L. (2012). Long-term differences in extinction risk among the seven forms of rarity. *Proc. R. Soc. B*, 279, 4969–4976.

Harnik, P.G., Fitzgerald, P.C., Payne, J.L. & Carlson, S.J. (2014). Phylogenetic signal in extinction selectivity in Devonian terebratulide brachiopods. *Paleobiology*, 40, 675–692.

Hirschfeld, M. & Rödel, M.-O. (2017). What makes a successful species? Traits facilitating survival in altered tropical forests. *BMC Ecol.*, 17, 25.

Hoffmann, M., Brooks, T., da Fonseca, G., Gascon, C., Hawkins, A., James, R. *et al.* (2008). Conservation planning and the IUCN Red List. *Endanger. Species Res.*, 6, 113–125.

Hopkins, M.J. & Klingenberg, C. (2011). How species longevity, intraspecific morphological variation, and geographic range size are related: a comparison using late Cambrian trilobites. *Evolution*, 65, 3253–3273.

Howard, S.D. & Bickford, D.P. (2014). Amphibians over the edge: silent extinction risk of Data Deficient species. *Divers. Distrib.*, 20, 837–846.

IUCN (2012). *IUCN Red List Categories and Criteria: Version 3.1*, 2nd edn. Gland, Switzerland; Cambridge, UK: IUCN. iv + 32pp.

IUCN (2017). Species extinction – the facts. Available at: https://www.iucn.org/sites/dev/files/import/downloads/species_extinction_05_2007.pdf . Last accessed 20 November 2017.

Jablonski, D. (2005). Mass extinctions and macroevolution. *Paleontol. Soc.*, 31, 192–210.

Jetz, W. & Freckleton, R.P. (2015). Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. *Philos. Trans. R. Soc. B*, 370, 20140016.

Kidwell, S.M. & Holland, S.M. (2002). The quality of the fossil record: implications for evolutionary analyses. *Annu. Rev. Ecol. Sys.*, 33, 561–588.

Kowalewski, M. (1996). Time-averaging, overcompleteness, and the geological record. *J. Geol.*, 104, 317–326.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A. *et al.* (2017). caret: Classification and regression training. R package version 6.0-76. Available at https://CRAN.R-project.org/package = caret.

Lindsey, C.C. (1966). Body sizes of poikilotherm vertebrates at different latitudes. *Evolution*, 20, 456–465.

Lips, K.R., Reeve, J.D. & Witters, L.R. (2003). Ecological traits predicting amphibian population declines in Central America. *Conserv. Biol.*, 17, 1078–1088.

Mace, G., Collar, N., Cooke, J., Gaston, K., Ginsberg, J., Leader Williams, N. *et al.* (1992). The development of new criteria for listing species on the IUCN Red List. *Species*, 19, 16–22.

Morais, A.R., Siqueira, M.N., Lemes, P., Maciel, N.M., De Marco, P. & Brito, D. (2013). Unraveling the conservation status of Data Deficient species. *Biol. Conserv.*, 166, 98–102.

Newell, N.D. (1959). The nature of the fossil record. *Proc. Am. Philos. Soc.*, 103, 264–285.

Nychka, D., Furrer, R., Paige, J. & Sain, S. (2015). fields: tools for spatial data. R package version 8.15. Available at: www.image.ucar.edu/fields.

Payne, J.L. & Finnegan, S. (2007). The effect of geographic range on extinction risk during background and mass extinction. *Proc. Natl Acad. Sci. U. S. A.*, 104, 10506–10511.

Purvis, A., Gittleman, J.L., Cowlishaw, G. & Mace, G.M. (2000). Predicting extinction risk in declining species. *Proc. R. Soc. B*, 267, 1947–1952.

Roger, B., Keitt, T. & Rowlingson, B. (2017). rgdal: bindings for the geospatial data abstraction library. R package version 1.2-8. Available at: https://CRAN.R-project.org/package = rgdal.

Rouse, J.D., Bishop, C.A. & Struger, J. (1999). Nitrogen pollution: an assessment of its threat to amphibian survival. *Environ. Health Perspect.*, 107, 799–803.

Ruland, F. & Jeschke, J.M. (2016). Threat-dependent traits of endangered frogs. *Biol. Conserv.*, 206, 310–313.

Sepkoski, J.J., Bambach, R.K., Raup, D.M. & Valentine, J.W. (1981). Phanerozoic marine diversity and the fossil record. *Nature*, 293, 435–437.

Signor, P.W. & Lipps, J.H. (1982). Gradual extinction patterns and catastrophes in the fossil record. *Geol. Soc. Am. Spec. Pap.*, 190, 291–296.

Sodhi, N.S., Bickford, D., Diesmos, A.C., Lee, T.M., Koh, L.P., Brook, B.W. *et al.* (2008). Measuring the meltdown: drivers of global amphibian extinction and decline. *PLoS ONE*, 3, e1636.

Tietje, M. & Rödel, M.-O. (2017a). Contradicting habitat type-extinction risk relationships between living and fossil amphibians. *R. Soc. Open Sci.*, 4, 170051.

Tietje, M. & Rödel, M.-O. (2017b). Data from: Evaluating the predicted extinction risk of living amphibian species with the fossil record. *GitHub Digital Repository*. Available at: https://doi.org/10.5281/zenodo.1206123.

Trochet, A., Moulherat, S., Calvez, O., Stevens, V.M., Clobert, J. & Schmeller, D.S. (2014). A database of life-history traits of European amphibians. *Biodivers. Data J.*, 2, e4123.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). Multivariate imputation by chained equations. *J. Stat. Softw.*, 45, 1–67.

Williams, S.E. & Hero, J.-M. (1998). Rainforest frogs of the Australian wet tropics: guild classification and the ecological similarity of declining species. *Proc. R. Soc. B*, 265, 597–602.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information section at the end of the article.

Editor, John Williams
Manuscript received 27 November 2017
First decision made 6 January 2018
Second decision made 17 March 2018
Manuscript accepted 23 March 2018

Paper 3

-

Climate influences the effect of range size on extinction risk in the amphibian fossil record

Tietje, M., Schobben, M., Rödel, M.-O. (manuscript in prep.) Climate influences the effect of range size on extinction risk in the amphibian fossil record.

Climate influences the effect of range size on extinction risk in the amphibian fossil record

Melanie Tietje[1], Mark-Oliver Rödel[1,2], Martin Schobben[3]

The geographic range size of many fossil and extant species has been shown to be of crucial importance for their extinction risk. This applies to amphibians as well, however, little is known about whether this influence varies, i.e. due to differing climatological conditions. We herein combine the amphibian fossil record with oxygen isotope data to test for the influence of climate on geographic range size, and on the correlation of range size and extinction risk. We show that geographic range size of amphibians and mean global temperature are not related, whereas the meridional temperature gradient is connected to both geographic range size, and the influence of range size on species durations.

Among the several factors that influence the extinction risk of species, the geographic range size has shown to be of strong importance both in studies on living and in the deep time fossil record for a large variety of different taxa (Finnegan et al., 2015; Kiessling and Aberhan, 2007; Kiessling and Kocsis, 2016; Harnik, 2011; Purvis et al., 2000; Sodhi et al., 2008; Tietje and Rödel, 2018). Based on this general pattern, geographic range size is also used as the main factor to assess the extinction risk for living species within the IUCN Red List assessment procedure (IUCN, 2012). The effect of range size on extinction risk has been assessed on different taxonomical and temporal scales, including the effect of range size changes (Kiessling and Kocsis, 2016). The influence of range size on extinction risk seems stable across various scales; however, temporal changes have not yet been assessed to our knowledge. We chose amphibians, one of the most threatened groups of terrestrial vertebrates (Stuart et al., 2008), which as ectotherms may experience a strong influence of temperature on geographic range size, which again affects extinction risk (Whitton et al., 2012). Here we assess the combination of geographic range size of extinct amphibian species and climate data from the geological record. We expect this to yield important insights in the interplay of range size, temperature, and the effect of range size on extinction risk under differing climatic conditions.

Material and Methods

Data

Occurrences of 354 extinct amphibian species were downloaded from the Paleobiology Database (PBDB) with no restrictions on time or area. The data were downloaded on 19 August 2014 via `http://fossilworks.org` (Alroy, 2016) and updated with a newer version on 15 February 2016. Downloaded data included information on species stratigraphic range (duration) and geographic range size (paleocoordinates). The dataset covers amphibian species from all Lissamphibia plus Temnospondyli and Lepospondyli, to account for the temnospondyl- and lepospondyl hypothesis of 'modern' amphibian origin, respectively (Schoch, 2014). Species duration was defined as the time between the midpoints of the youngest and oldest stratigraphic unit in which a species occurred, rounded to the next full million years. The stratigraphic units of our data is chronostratigraphic stages as provided by the PBDB. Geographic range size was calculated as the great circle distance between a species occurrences for each geologic stage. The final values for the species are the maximum value of all geologic stages; and maximum great circle distances (see Tietje and Rödel, 2018). As temperature proxy for the Phanerozoic we used the oxygen isotope data (δ^{18}O measurements) provided by Veizer and Prokoph (2015) and reduced this dataset to brachiopod, belemnite, bivalve and planktonic foraminiferal δ^{18}O data. Each isotope value is referenced to the geologic stage (GTS 2012) and comes with information on the climatic zone (temperate, arctic, tropical, subtropical). Isotope values were corrected for diagenetic δ^{18}O depletion (phanerozoic trend, "secular shift") using the equation provided by Veizer and Prokoph (2015).

Methods

To cancel out the uneven spread of data in the isotope dataset a moving window approach was applied and the medians of each window calculated (Loader, 1999). Additionally, we applied a subsampling routine; within each window the data were subsampled (n=100) to the size of the minimum number of isotope values contained in all windows, the final value being the median of all subsamples. The global mean relative temperature was calculated using the tropical δ^{18}O data from Veizer and Prokoph (2015). For estimating the meridional temperature gradient, we restricted the dataset to the occurrences of planctonic foraminifera to avoid increased isotope fractionation effects in high-latitude brachiopod data. The isotope data were split into two sets according to the given climatic zone information. "Arctic" and "temperate" labelled data points were used to calculate the high latitude temperature values, data points labelled with "tropical" were used to calculate low latitude temperature values. The moving window approach, including subsampling as described above, was applied to both datasets. The temperature gradient was calculated as the difference between the two resulting median values from both climatic zones. Data labelled as subtropical were not used for the temperature gradient due to potentially high evaporation in these areas, which can lead to falsely increased δ^{18}O values, known as the cool tropic paradox (D'Hondt and Arthur, 1996).

To estimate changes in the strength of the association of geographic range size and species' duration we calculated their correlation (Spearman rank correlation) for every geologic period from the Carboniferous to the Neogene. To account for sample size differences between geologic periods, we additionally performed the correlation analysis on bootstrapped data (repeated sampling (500 repetitions)

[1]Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity, D-10115 Berlin, Germany [3] Universiteit Utrecht, Utrecht, The Netherlands [2] Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-10115 Berlin, Germany

with the sample size per period being the smallest number of data points in all periods (n=19)), resulting in median correlation values and medium absolute deviation as an error measure. To obtain a measure for the possible variation of the overall correlation between range size and duration (due to sampling etc), we repeatedly sampled the data regardless of the time frame and applied correlation analysis on these subsamples (200 repetitions, n=30). We then tested if variations in the correlation of range size with duration were linked to temperature or temperature gradient. Correlations were calculated for geologic periods, and for 3 to 7 equally sized time bins, ranging from 52 to 16 million years. Time-series data were tested for autocorrelation before performing correlation tests. There were no significant autocorrelations in any of the relevant variables (great circle distance, global mean relative temperature, global temperature gradient).

All analysis was done in the R environment version 3.4.4. Data and R code to replicate this study are fully available on [github repository DOI].

Results

Correlation of geographic range size and species' duration

The correlation of geographic range size with the duration of species was stable over time (figure 1). Slight variations were within the two-times standard deviation of the data and thus not significant.

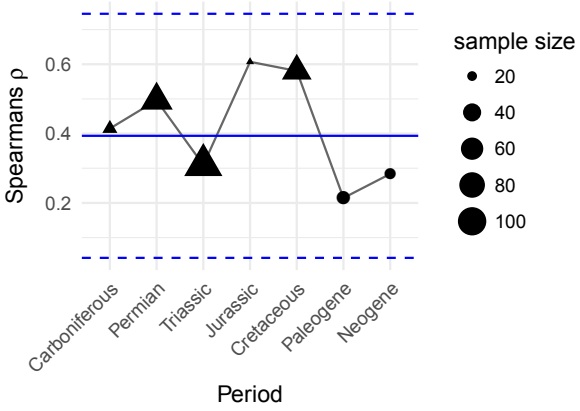

Figure 1 – Correlation of geographic range size and duration of species in different geologic periods (Spearman's rank test). The number of species per period is visualized as dot size, significant correlation values (p < 0.05) are displayed as triangles, insignificant correlation values depicted as circles. The solid blue line depicts the overall mean correlation, the dashed lines are the two-times standard deviation of the correlation distribution gained from repeated calculation of the correlation from a random subset of the data (200 repetitions).

Correlation of geographic range size and climate

Geographic range size was not correlated to the mean relative temperature, but negatively to the temperature gradient (Spearman's rho= -0.33, $p < 0.01$). This correlation has shown to be stable in cross-validation via bootstrap (figure S4).

Climatic influence on the geographic range size - duration correlation

The correlation strength of geographic range size with duration was not linked to the mean relative temperature, but showed a trend towards a strong, negative correlation with the temperature gradient (figure 2). This trend is negative for all different bin sizes, however only significant for bin sizes 5 and 7 (slope = -0.36 and -0.32, p = 0.02 and 0.04, respectively), probably a result of the low number of time bins.

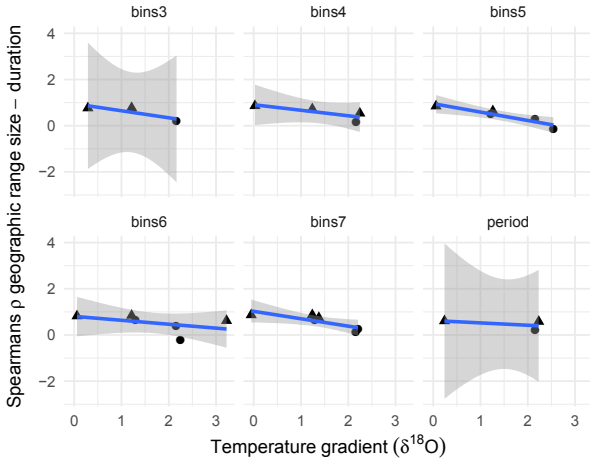

Figure 2 – Temporal changes in correlation strength of range size and species' duration, measured as Spearmans rho. Significant correlation values (p < 0.05) are displayed as triangles, insignificant correlation values depicted as circles. Correlations are shown for data divided in geologic periods and in 3 to 7 equally sized time bins. Blue lines depict linear models, shaded areas the confidence interval of the model.

Table 1 – Summaries of the linear models depicted in figure 2.

Time bins	Adj. R^2	Slope	p-value
3	0.49	-0.30	0.33
4	0.48	-0.24	0.19
5	0.84	-0.36	**0.02**
6	0.02	-0.17	0.35
7	0.59	-0.32	**0.04**
Periods (3)	-0.45	-0.10	0.65

Discussion

We have shown the temporal stability of the correlation of geographic range size with species' duration. We also found the correlation strength likely being negatively correlated with the meridional temperature gradient, suggesting a decreasing importance of the geographic range size for duration in a more diverse world.

We detected that the connection between species' range sizes and durations was temporally stable. This temporal stability supports the ubiquitous influence of this trait under differing environmental conditions. It also supports the important role of the geographic range size for the estimation of extinction risk for extant species, as it is included in the current assessment procedures of the IUCN Red List (IUCN, 2012). The temporal stability also encourages the utilization of the fossil record for modeling the extinction risk of living species (Finnegan et al., 2015; Tietje and Rödel, 2018).

The minor changes that we did observe in the temporal correlation pattern (figure 1), as well as the geographic range size itself, were negatively correlated with the global temperature gradient. We interpret a large temperature gradient as a broader scale of differing suitable potential habitats. Besides natural dispersal barriers like mountain ranges or oceans, the latitudinal temperature gradient adds another limitation to the (at least latitudinal) spread of a species. Major terrestrial biomes are mainly coupled to the mean annual temperature and precipitation, which are both a result of the latitudinal climatic zonation (Cox and Moore, 1993; Zubakov and Borzenkova, 1990).

Our results suggest that a large temperature gradient shrinks the latitudinal boundaries for geographic range sizes and therefore leads to smaller geographic ranges in general. Given that amphibians as ectotherms highly depend on suitable temperatures to function and survive (Wells, 2007), this connection is not unexpected.

Our results also suggest that an increasing temperature gradient is associated with a weakening of the range size-duration connection, which can be interpreted as the importance of range size for a species extinction risk being highest when the potential habitat is largest. Vice versa, a small range size in a large temperature gradient world would not be as detrimental as a similar range size in a more homogenous world, because range size seems less important under these circumstances. One possible reason could be that smaller suitable areas result in higher levels of interspecific competition due to the shrinking available space. Therefore, other factors besides the geographic range size might become more important for survival, like the ability to cope with competitors and exploit limited resources. However, limited space could also lead to stronger niche partitioning instead of higher levels of interspecific competition. Therefore, depending on the time after which niche partitioning starts to become effective, the increased interspecific competition might be most pronounced in times of temperature gradient changes. Unfortunately, our dataset does not provide the temporal resolution to observe these fine-scale patterns. On the other end, limited competitive qualities might be compensated with a large geographic range size, raising the importance of this trait for extinction risk as soon as the environmental conditions allow for a wider spread.

Another factor might be that a general upper range size limit, caused by environmental restrictions, does not allow for geographic range sizes large enough to strongly influence the duration. This would suggest however a non-linear relationship between range size and duration that is mainly driven by large ranges. Differences between large ranges would be more important than differences between small ranges.

Temporal variations in correlation strength between rang size and extinction risk might be explained by changing climatic conditions. However these variations were minor and do not questions the strong influence of range size on extinction risk. The stability of this trait supports its suitability for its usage in the assessment of extinction risk, marking a step forward in the credibility of big scale paleobiological data analysis contributing to conservation biology tasks.

References

Alroy, J. Fossilworks. Gateway to the Paleobiology Database. http://fossilworks.org, 2016. URL `http://fossilworks.org`.

Cox, C. B. and Moore, P. D. *Biogeography and Ecology*. Blackwell Science Ltd, London, UK, 5 edition, 1993.

D'Hondt, S. and Arthur, M. A. 1996. Late Cretaceous oceans and the Cool Tropic Paradox. *Science*, 271(5257): 1838–1841.

Finnegan, S., Anderson, S. C., Harnik, P. G., Simpson, C., Tittensor, D. P., Byrnes, J. E., Finkel, Z. V., Lindberg, D. R., Liow, L. H., Lockwood, R., Lotze, H. K., McClain, C. R., McGuire, J. L., and Pandolfi, J. M. 2015. Paleontological baselines for evaluating extinction risk in the modern oceans. *Science*, 348(6234): 567–570.

Harnik, P. G. 2011. Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33): 13594–13599.

IUCN. *IUCN Red List Categories and Criteria: Version 3.1*. Gland, Switzerland and Cambridge, UK: IUCN. iv + 32pp, second edi edition, 2012.

Kiessling, W. and Aberhan, M. 2007. Geographical distribution and extinction risk: lessons from Triassic-Jurassic marine benthic organisms. *Journal of Biogeography*, 34 (9): 1473–1489.

Kiessling, W. and Kocsis, Á. T. 2016. Adding fossil occupancy trajectories to the assessment of modern extinction risk. *Biology Letters*, 12(10): 20150813.

Loader, C. *Statistics and Computing*, volume 42. Springer-Verlag, New York Berlin Heidelberg, 1999.

Purvis, A., Gittleman, J. L., Cowlishaw, G., and Mace, G. M. 2000. Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, 267: 1947–1952.

Schoch, R. R. *Amphibian evolution: the life of early land vertebrates*. Number April 2012. Wiley-Blackwell, UK, 2014.

Sodhi, N. S., Bickford, D., Diesmos, A. C., Lee, T. M., Koh, L. P., Brook, B. W., Sekercioglu, C. H., and Bradshaw, C. J. A. 2008. Measuring the meltdown: drivers of global amphibian extinction and decline. *PLoS ONE*, 3(2): e1636.

Stuart, S., Hoffmann, M., Chanson, J., Cox, N., Berridge, R., Ramani, P., and Young, B., editors. *Threatened Amphibians of the World.* Lynx Edicions, Barcelona, Spain; IUCN, Gland, Switzerland; and Conservation International, Arlington, Virginia, USA, 2008.

Tietje, M. and Rödel, M.-O. 2018. Evaluating the predicted extinction risk of living amphibian species with the fossil record. *Ecology Letters*, 21(8): 1135–1142.

Veizer, J. and Prokoph, A. 2015. Temperatures and oxygen isotopic composition of Phanerozoic oceans. *Earth-Science Reviews*, 146: 92–104.

Wells, K. D. Temperature Relations. In *The ecology and behavior of amphibians*, chapter 3, pages 122–156. The University of Chicago Press, Chicago and London, 2007.

Whitton, F. J. S., Purvis, A., Orme, C. D. L., and Olalla-Tárraga, M. Á. 2012. Understanding global patterns in amphibian geographic range size: does Rapoport rule? *Global Ecology and Biogeography*, 21(2): 179–190.

Zubakov, V. A. and Borzenkova, I. I. Global Paleoclimate of the Late Cenozoic. In *Developments in Palaeontology and Stratigraphy*, page 456. Elsevier, Amsterdam, 1990.

Acknowledgements This project was funded by the Museum für Naturkunde. We thank the numerous contributours to the Paleobiology Database for entering their data and making quantitative studies possible. This is Paleobiology database publication number XXX.

6 Supplementary material

Supplementary material: Paper 1

Contradicting habitat type-extinction risk relationships between living and fossil amphibians

Contradicting habitat type - extinction risk relationships between living and fossil amphibians

Melanie Tietje and Mark-Oliver Rödel

Supplement

Data files

- R script and data files: *Tietje_Rodel_Rscript_and_data_files_ESM.tar*
- Paleobiology Database download reference data: *Tietje_Rodel_pbdb_references_ESM.csv*
- Lithology and specimen completeness metric references: *Tietje_Rodel_lithology_and_SCM_references_ESM.csv*

Tables

Table S1 Usage of taxonomic group names in the databases FosFAR (Database of Vertebrates: Fossil Fishes, Amphibians, Reptiles, Birds) and Paleobiology Database (PbDb) in this publication.

Taxon name used in this publication	FosFARbase	PbDb	Description
Allocaudata	Allocaudata Fox&Naylor, 1982	Albanerpetontidae Fox&Naylor, 1982	The included species are identical. PbDb does not use Allocaudata Fox & Naylor, 1982 as order though.
Species from stem groups were assigned to their corresponding pan-groups Salientia, Urodela and Parabatrachia	-	Amphibia Linnaeus, 1758	Amphibia covers all stem group taxa in amphibians that were not covered by Anura, Caudata or Gymnophiona, it also includes Lepospondyli. This taxonomic rank is not available in FosFARbase.
Salientia. Includes Anura and their stem-taxa [1]	Anura Rafinesque, 1815	Anura Fischer von Waldheim, 1813	PbDb Anura contains crown-group Anura, FosFARbase Anura contains stem-group Anura like *Triadobatrachus massinoti*
Urodela. Includes all Caudata and their stem-taxa [1]	Caudata Scopoli, 1777	Caudata Scopoli, 1777	FosFARbase includes stem-taxa only, PbDb includes both stem- and crown-taxa.
Parabatrachia. Includes Gymnophiona and their stem-taxa [1]	Gymnophiona Rafinesque, 1814	Gymnophiona Rafinesque, 1814	Both databases include stem group taxa despite using the definition by Rafinesque 1814.
Lepospondyli	-	Lepospondyli, von Zittel 1887	Lepospondyli were only available from PbDb, included in Amphibia
Temnospondyli. Without Lissamphibia[2]	Temnospondyli Von Zittel, 1888	Temnospondyli von Zittel, 1887	-
Urodela. Includes Caudata and their stem-taxa [1]	Urodela Latreille, 1825	Urodela Duméril, 1805	Groups in both databases contain the crown-group Caudata.

Table S2 Species name synonyms. We used PbDb names in our analysis.

PbDb name	fosFAR name	Reference
Albanerpeton nexuosus	*Albanerpeton nexuosum*	[3]
Albanerpeton pannonicus	*Albanerpeton pannonicum*	[3]
Rewana myriadens	*Arcadia myriadens*	[3]
Koskinonodon perfectus	*Buettneria howardensis*	[3]
Koskinonodon perfectus	*Buettneria perfecta*	[3]
Gerrothorax pulcherrimus	*Gerrothorax rhaeticus*	[3]
Tatrasuchus wildi	*Kupferzellia wildi*	[3]
Palaeobatrachus tobieni	*Messelobatrachus tobieni*	[3]
Meyerosuchus fuerstenbergianus	*Meyerosuchus fuerstenberganus*	[3]
Scaphiopus alexanderi	*Spea alexanderi*	[3]
Spea pliobatrachus	*Spea pliobatracha*	[3]
Bufo defensor	*Anaxyrus defensor*	[4]
Bufo hibbardi	*Anaxyrus hibbardi*	
Bufo repentinus	*Anaxyrus repentinus*	
Bufo rexroadensis	*Anaxyrus rexroadensis*	
Bufo tiheni	*Anaxyrus tiheni*	
Bufo valentinensis	*Anaxyrus valentinensis*	

Table S3 Completeness measure estimates for different taxonomic levels in amphibians. “No REP” indicates no range end points of species were included in the calculations.

Group	**Proportion**
Simple completeness metric (SCM)	
Species	0.94
Species no REP	0.44
Genera	0.78
Genera no REP	0.35
Families	0.60
Families no REP	0.28
Families Cretaceous	0.43
Families Cretaceous no REP	0.30
Preservation probability based on duration frequency	
Species	0.35
Genera	0.47
Proportion of living taxa with fossil record	
Families	0.33
Genera	0.08
Species	0.008
Proportion of living taxa with fossil record for mammals	
Families	0.69
Genera	0.03
Species	0.002

Table S4 Comparisons of specimen completeness in different habitat categories (level 1). Groups were *stagnant* (81), *low-velocity* (72), *medium-velocity* (35) and *high-velocity* (14). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (P-value adjustment: fdr). Significant p-values ($p < 0.05$) are shown in bold.

	Groups		
Kruskal-Wallis test			
χ^2 (df)	18.03 (3)		
p-value	**0.0004**		
Wilcoxon rank-sum test			
	Stagnant	*Low-velocity*	*Medium-velocity*
Low-velocity	**0.0013**	-	-
Medium-velocity	**0.0222**	0.8956	-
High-velocity	0.1369	0. 8956	0. 8956

Table S5 Comparisons of specimen completeness in different taxonomic groups. We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (P-value adjustment: fdr). Significant p-values ($p < 0.05$) are shown in bold. Group sizes were: No-Lepospondyli (189), No-Temnospondyli (159), Lepospondyli (13), Temnospondyli (43).

	Groups		
Kruskal-Wallis test	*Lissamphibia*	*Others*	
χ^2 (df)	6.54 (3)	45.93 (3)	
p-value	0.09	**< 0.0001**	
Wilcoxon rank-sum test			
	No-Lepospondyli	No-Temnospondyli	Lepospondyli
No-Temnospondyli	0.27	-	-
Lepospondyli	**< 0.0001**	**< 0.0001**	-
Temnospondyli	**< 0.0001**	**< 0.0001**	0.10

Table S 6 Comparison of species duration from different environments (for category definitions see Fig. 1). Groups were *stagnant* (214), *low-velocity* (130), *medium-velocity* (56) and *high-velocity* (18); *lentic* (216) and *lotic* (176); *low* (319) and *high* (71). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (p-value adjustment: fdr). Significant differences ($p < 0.05$) are given in bold.

	Level 1	Level 2	Level 3
Kruskal-Wallis test			
χ^2 (df)	17.86 (3)	4.33 (1)	17.27 (1)
p-value	**0.0005**	**0.0375**	**<0.0001**
Wilcoxon rank-sum test for level 1			
	Stagnant	Low-velocity	Medium-velocity
Low-velocity	0.0568	-	-
Medium-velocity	**0.0001**	0.1128	-
High-velocity	0.2385	0.8187	0.5071

Table S7 Comparisons of species duration between different taxonomic groups. Groups were Allocaudata (11), Urodela (39), Parabatrachia (2), Salientia (80) and No-Temnospondyli (171), No-Lepospondyli (310), Lepospondyli (39), Temnospondyli (178). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (P-value adjustment: fdr). Significant p-values ($p < 0.05$) are shown in bold.

Lissamphibia group			
Kruskal-Wallis test			
χ^2 (df)	12.59 (3)		
p-value	**0.0056**		
Wilcoxon rank-sum test			
	Allocaudata	Parabatrachia	Salientia
Parabatrachia	0.3569	-	-
Salientia	**0.0059**	0.5358	-
Urodela	0.1717	0.4638	0.1374

Other groups			
Kruskal-Wallis test			
χ^2 (df)	9.31 (3)		
p-value	**0.0255**		
Wilcoxon rank-sum test			
	No - Lepospondyli	No - Temnospondyli	Lepospondyli
No –Temnospondyli	0.2640	-	-
Lepospondyli	0.3800	0.2100	-
Temnospondyli	0.1770	**0.0240**	0.9630

Table S8 Observed proportions divided by expected proportions of species occurrences in each environment and taxonomic group under the assumption of equal proportions distribution. Results were computed using Pearson's chi-square test. Deviations larger than 20% are shown in bold.

	Stagnant	Low-velocity	Medium-velocity	High-velocity
Others				
No-Lepospondyli	0.94	1.07	1.07	0.99
No-Temnospondyli	1.09	0.76	**1.27**	**0.77**
Lepospondyli	**1.57**	**0.32**	0.36	1.14
Temnospondyli	0.91	**1.24**	**0.72**	1.24

Table S9 Comparisons of durations of species from different environments and groups. We used Kruskal-Wallis rank sum test (χ^2 (df), p-value). Significant p-values ($p < 0.05$) are given in bold. Pairwise comparisons for groups with significant differences are displayed in Table S 10.

	Level 1		*Level 2*		*Level 3*	
Lissamphibia	p-value	χ^2(df)	p-value	χ^2 (df)	p-value	χ^2 (df)
Allocaudata	0.7861	1.1 (3)	0.7843	0.1 (1)	0.8616	0.03 (1)
Salientia	**0.0025**	14.3 (3)	**0.0082**	6.9 (1)	**0.0234**	5.14 (1)
Urodela	0.0818	5.0 (2)	0.1062	2.6 (1)	0.1017	2.7 (1)
Others						
No – Temnospondyli	**<0.0001**	14.2 (3)	**0.0014**	10.2 (1)	**0.0017**	9.9 (1)
No – Lepospondyli	**0.0028**	22.4 (3)	0.0649	3.4 (1)	**0.0013**	14.5 (1)
Lepospondyli	0.3673	3.4 (3)	0.9765	0	0.4559	0.6 (1)
Temnospondyli	0.5034	2.3 (3)	0.9754	0	0.0530	3.7 (1)

Table S10 P-values for pairwise comparisons of durations in level 1 habitat categories in the groups Salientia, No-Lepospondyli and No-Temnospondyli. We used Pairwise Wilcoxon Rank Sum Tests with fdr p-value adjustment). Significant p-values ($p < 0.05$) are shown in bold.

	No - Lepospondyli	No - Temnospondyli	Salientia
Stagnant / low-velocity	0.0859	**0.0002**	**0.0068**
Stagnant / medium-velocity	**0.0012**	**0.0002**	**0.0032**
Stagnant / high-velocity	0.2013	0.5459	0.7064
Low-velocity / medium-velocity	0.2013	0.9849	0.7064
Low-velocity / high-velocity	0.7021	0.5459	0.4837
Medium-velocity / high-velocity	0.7021	0.5459	0.4837

Table S11 Comparison of species geographic ranges from different environments (for category definitions see Fig. 1). Groups were *stagnant* (214), *low-velocity* (129), *medium-velocity* (56) and *high-velocity* (18); *lentic* (214) and *lotic* (175); *low* (318) and *high* (71). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (p-value adjustment: fdr). Significant differences ($p < 0.05$) are given in bold.

	Level 1	Level 2	Level 3
Kruskal-Wallis test			
χ^2 (df)	19.1 (3)	0.42 (1)	19.3 (1)
p-value	0.0003	0.5161	**< 0.0001**
Wilcoxon rank-sum test for level 1			
	Stagnant	Low-velocity	Medium-velocity
Low-velocity	0.5858	-	-
Medium-velocity	**0.0242**	0.0946	-
High-velocity	**0.0004**	**0.0027**	0.1519

Table S12 Comparison of species duration from different environments, controlled for geographic range. Groups were split according to the geographic range of the species being 1 or more grid cell counts. Groups for were *stagnant* (185|29), *low-velocity* (110|20), *medium-velocity* (41|15) and *high-velocity* (9|9); *lentic* (185|29) and *lotic* (148|28); *low* (284|35) and *high* (49|22). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (p-value adjustment: fdr). Significant differences ($p < 0.05$) are given in bold.

	Level 1		Level 2		Level 3	
	1	>1	1	>1	1	>1
Kruskal-Wallis test						
χ^2 (df)	14.58 (3)	1.67(3)	6.61 (1)	0.002(1)	6.51 (1)	0.72(1)
p-value	**0.0022**	0.6443	**0.0176**	0.967	**0.0107**	0.3958
Wilcoxon rank-sum test for level 1, range = 1						
	Stagnant	Low-velocity	Medium-velocity			
Low-velocity	0.0576	-	-			
Medium-velocity	**0.0022**	0.2244	-			
High-velocity	0.3791	0.2244	0.1413			

Table S13 Deviation of observed from expected proportions of single-interval species between energetic settings according to Pearson's chi-squared test.

Habitat	Single-interval	Longer
Stagnant	1.08	0.70
Low-velocity	0.97	1.11
Medium-velocity	0.78	1.87
High-velocity	0.97	1.11

Table S 14 Comparison of species duration from different environments excluding single-interval species (for category definitions see Fig. 1). Groups were *stagnant* (30), *low-velocity* (29), *medium-velocity* (21) and *high-velocity* (4); *lentic* (30) and *lotic* (38); *low* (42) and *high* (24). We used Kruskal-Wallis rank sum test (χ^2 (df), p-value) and Pairwise Wilcoxon Rank Sum Tests for pairwise comparisons (p-value adjustment: fdr). Significant differences ($p < 0.05$) are given in bold.

	Level 1	Level 2	Level 3
Kruskal-Wallis test			
χ^2 (df)	3.88 (3)	1.06 (1)	3.77 (1)
p-value	0.1808	0.3039	0.0523

Figures

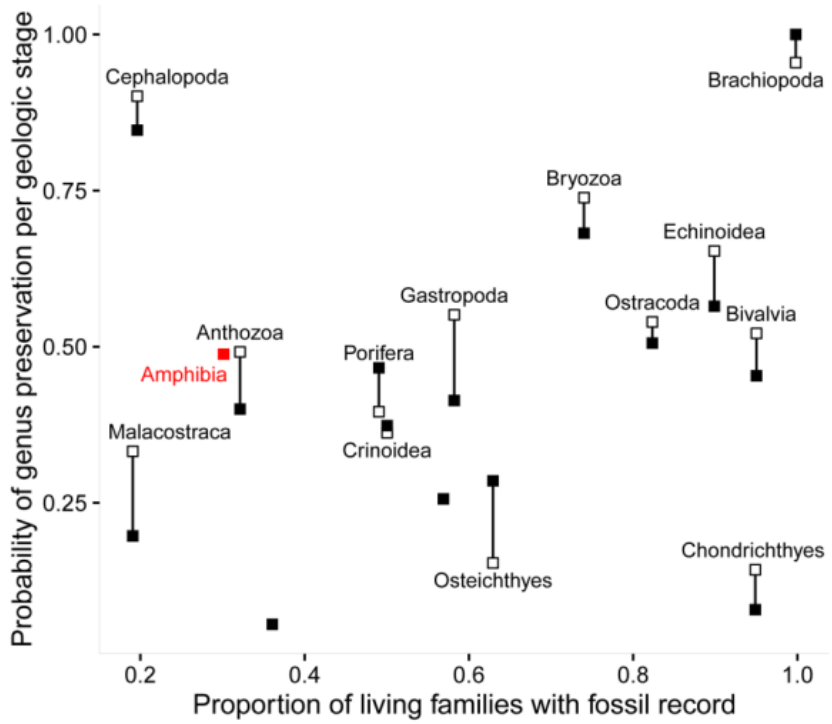

Figure S1 Proportion of living families with a fossil record plotted against probability of genus preservation (adapted from Figure 1 by Foote & Sepkoski [5]). The red dot shows values for our fossil amphibian dataset.

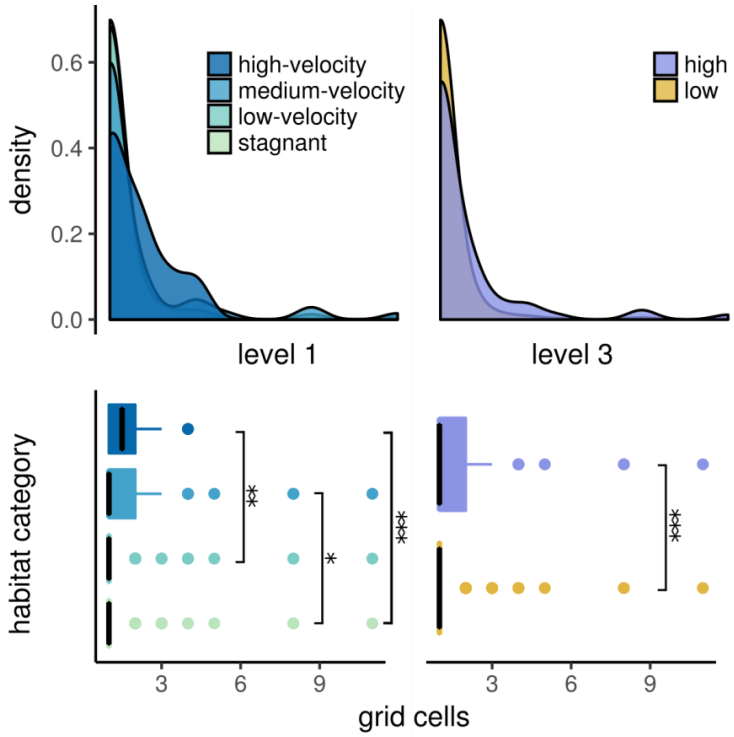

Figure S2 Geographic ranges in number of occupied grid cells for species in different environments. Categories and colour coding are as defined in Figure 1. Sample sizes for the groups are the same as in Figure 3. The upper panel shows the density distribution of durations (bandwidth=2 million years), the lower panel shows the durations as boxplots, with black lines indicating the median and coloured areas showing the range between first and third quartiles. Significance levels are indicated by one, two and three asterisks representing p-values smaller than 0.05, 0.01 and 0.001, respectively.

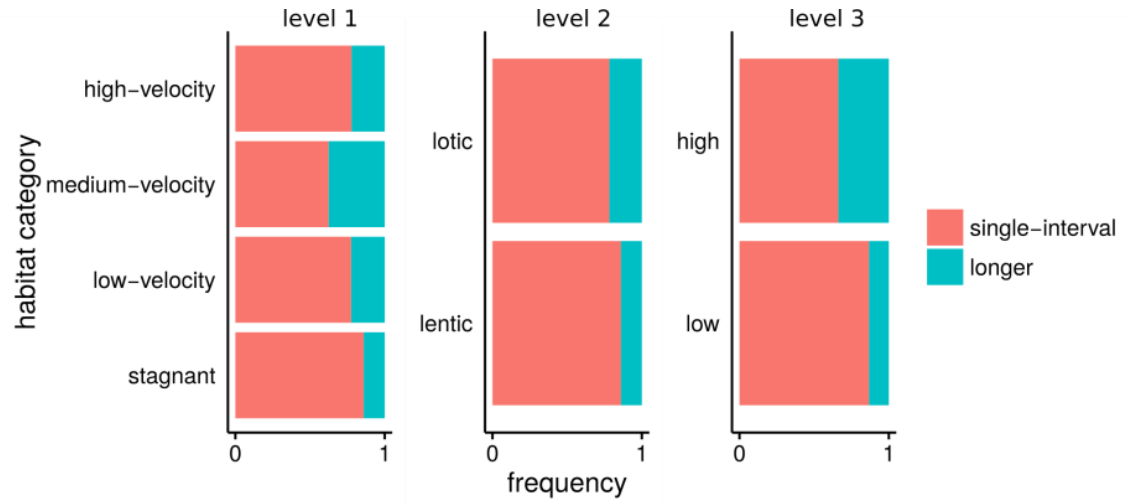

Figure S3 Frequencies of single-interval species in different habitat categories.

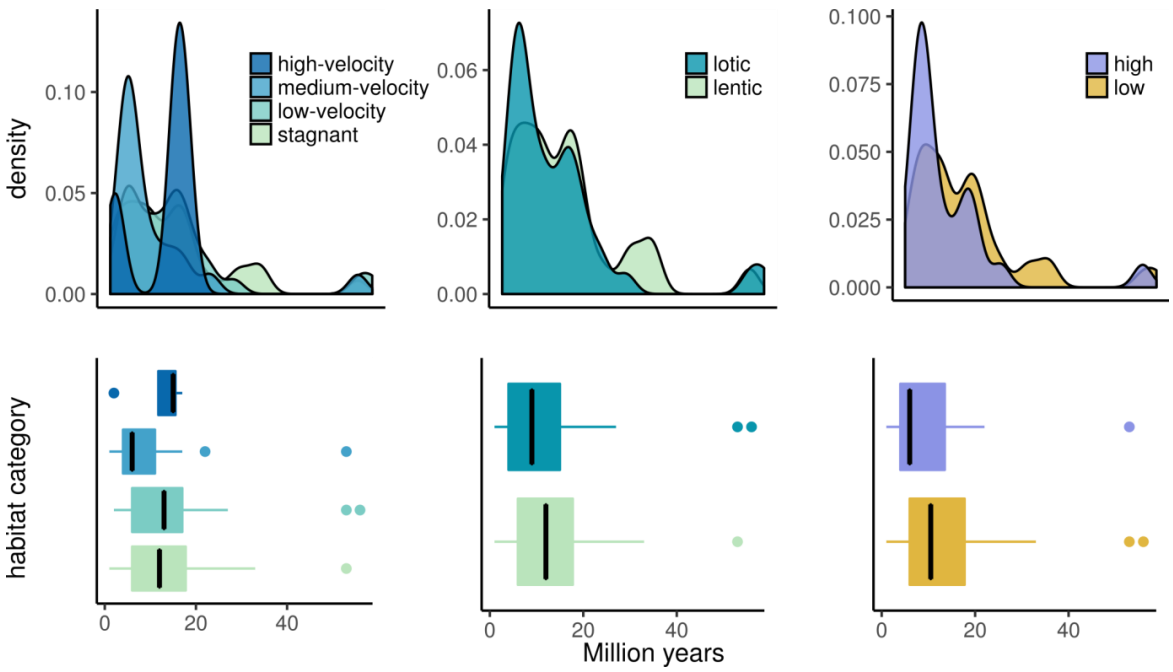

Figure S4 Durations of amphibian species excluding single-interval species in different environments. Species were grouped into four basic (level 1) and two broader environmental categories (level 2 and 3; compare Figure 1 in the manuscript. Sample sizes for groups were: stagnant (30), low-velocity (29), medium-velocity (21), high-velocity (4); lentic (30) and lotic (38); low (42) and high (24).

References supplement

1. Frost DR et al. 2006 The Amphibian Tree of Life. *Bull. Am. Museum Nat. Hist.* **297**, 1–291. (doi:10.1206/0003-0090(2006)297[0001:TATOL]2.0.CO;2)
2. Schoch RR, Milner AR. 2014 Handbook of Paleoherpetology - Part 3A2: Temnospondyli I. Verlag Dr. Friedrich Pfeil, München
3. Fossilworks 2016. http://fossilworks.org
4. Martín C, Alonso-Zarazaga MA, Sanchiz B. 2012 Nomenclatural notes on living and fossil amphibians. *Graellsia* **68**, 159–180. (doi:10.3989/graellsia.2012.v68.056)
5. Foote M, Raup DM. 1996 Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* **22**, 121–140. (doi:10.1017/S0094837300016134)

Supplementary material: Paper 2

Evaluating the predicted extinction risk of living amphibian species with the fossil record

Appendix S1 Evaluating the predicted extinction risk of living amphibian species with the fossil record

Supplementary figures, tables and modelling output

Melanie Tietje (Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany)
Mark-Oliver Rödel (Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Germany; Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Germany)

Contents

Intro

This document contains supplementary figures S1-20, tables S1-6, model output files and results from Kruskal-Wallis rank sum tests as well as Pairwise Wilcoxon rank sum tests. It is writen with the knitr package (Xie 2014, 2015, 2016) for the R statistical environment (R Core Team 2017). The entire analysis can be recreated by running knitr::knit() on the Markdown file *ELEtietjeSA1.Rmd*. The necessary files are available here (http://doi.org/10.5281/zenodo 1206123) and in the Git repository: https://github.com/Eryops1/supplement_amphibian_extinction_risk.

We used R version 3.4.3 (2017-11-30) and the following packages:

```
loadedNamespaces()
```

```
##  [1] "maps"         "ddalpha"      "tidyr"        "sfsmisc"
##  [5] "splines"      "foreach"      "gsubfn"       "prodlim"
##  [9] "dotCall64"    "assertthat"   "stats4"       "sp"
## [13] "grDevices"    "DRR"          "yaml"         "robustbase"
## [17] "ipred"        "pillar"       "backports"    "lattice"
## [21] "glue"         "base"         "digest"       "randomForest"
## [25] "colorspace"   "recipes"      "gbm"          "captioner"
## [29] "htmltools"    "Matrix"       "plyr"         "psych"
## [33] "timeDate"     "pkgconfig"    "CVST"         "broom"
## [37] "caret"        "purrr"        "scales"       "gower"
## [41] "lava"         "tibble"       "mgcv"         "datasets"
## [45] "ggplot2"      "withr"        "nnet"         "lazyeval"
## [49] "mnormt"       "proto"        "survival"     "magrittr"
## [53] "evaluate"     "methods"      "mice"         "nlme"
## [57] "MASS"         "dimRed"       "foreign"      "utils"
## [61] "class"        "tools"        "stringr"      "kernlab"
## [65] "munsell"      "bindrcpp"     "stats"        "compiler"
## [69] "RcppRoll"     "rlang"        "grid"         "simpleboot"
## [73] "iterators"    "graphics"     "spam"         "tcltk"
## [77] "rmarkdown"    "boot"         "gtable"       "ModelMetrics"
## [81] "codetools"    "reshape"      "reshape2"     "R6"
## [85] "lubridate"    "gridExtra"    "knitr"        "dplyr"
## [89] "rgdal"        "bindr"        "rprojroot"    "stringi"
## [93] "parallel"     "Rcpp"         "fields"       "rpart"
## [97] "tidyselect"   "DEoptimR"
```

Data collection

Extinct species data

Table S1: Variable description for extinct species.

Variable	Description
Duration	Duration of a species in the fossil record in million years. Time between the midranges of the oldest and youngest chronostratigraphic stage the species was observed, rounded to the next million years.
Abundance	Four categories for abundance were build based on the minimum number of individuals (MNI) and specimen counts for each species. MNI and specimen count data were obtained from the PBDB or literature. Maximum values were calculated per locality and stage. The final values for each species represent the maximum value ever shown by one species over space and time. These maximum values per species were clustered via k-means clustering into four numeric, ordinal categories (1, 2, 3 and 4). K-means clustering is a 1-n dimensional approach which tries to minimize the sum of squares within each cluster, that is the distance between the points and the center of each cluster. This is accomplished by repeatedly and randomly setting the position of the centers in the 2-dimensional space, searching for the optimal solution (Harting & Wong 1979). MNI and specimen counts are the two dimensions used in this clustering.
Geographic range	Geographic range size is calculatd as maximum great circle distance (shortest connection between two points on the surface of a sphere). Great circle distances were calculated for each species in each stage. The maximum values ever achieved by a species is used.
Latitudinal range	Maximum latitudinal range of a species (maximum difference between paleolatitudes). Ranges were calculated for each species in each stage. The maximum values ever achieved is used.
Mean latitude	Mean latitude calculated from all fossil occurrence coordinates of each species.
Minimum latitude	Minimum latitude calculated from all fossil occurrence coordinates of each species.
Body size	Maximum snout-vent-length (SVL) of the species. We collected SVL whenever possible, if not available we collected total length (TL) and absolute skull length (ASL). To estimate the SVL from TL and ASL, we created linear regressions that connect SVL, TL and ASL for species where all or at least two measurements were available. These linear models were used to calculate the SVL for species from TL or ASL. When measures on body size where not available, we used the bodysize of a congeneric species.

Stratigraphic resolution

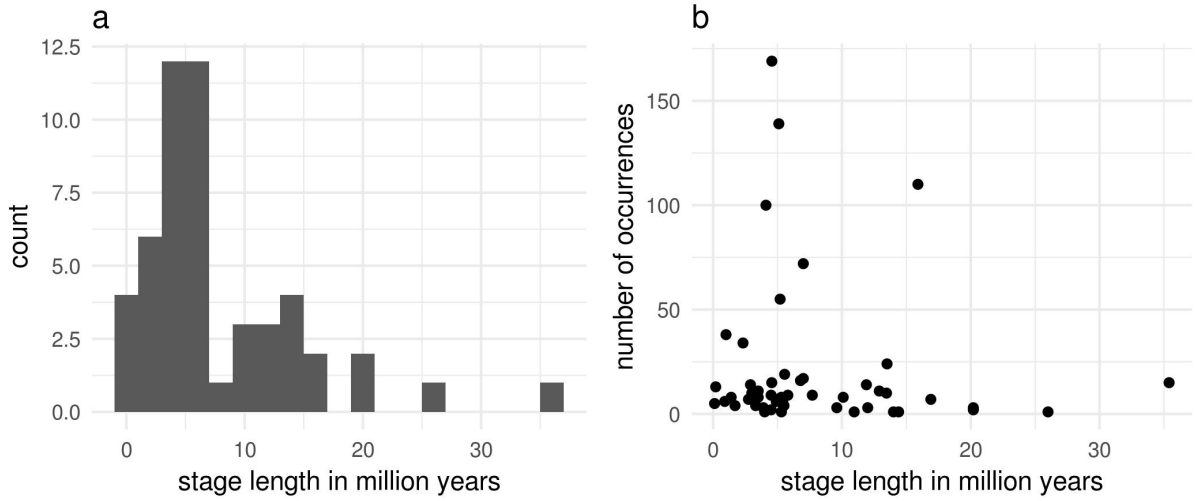

Figure S1: Stratigraphic resolution of the fossil data. Panel a) shows a histogram of the stage lengths (n=51). Panel b) shows the number of occurrences in each stage.

Extinct species data summary

Table S2: Summary statistics for the number of occurrences per species.

Variable	Value
Average number of occurrences per species	2.44
Minimum number of occurrences per species	1
Maximum number of occurrences per species	116

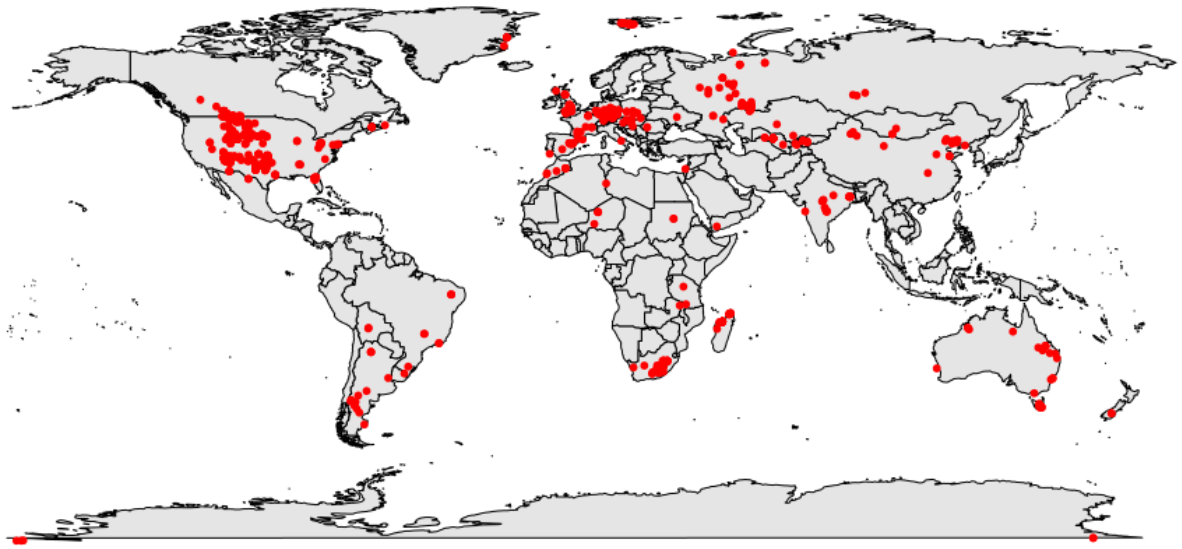

Figure S2: A map of all amphibian fossil occurrences used in the analysis.

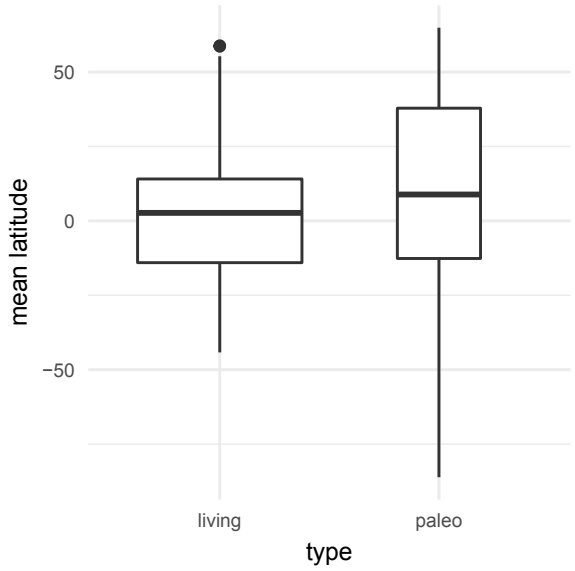

Figure S3: Boxplot showing the mean latitude of each living and fossil species. Boxplot width is scaled to sample size of each group (living: 1382, paleo: 354).

Living species data

Abundance data

Table S3: Keywords for text mining the abundance information of the IUCN Red List webpage for living amphibian species.

Category	Keywords	Conditioning
1	Not common, fewer than, rare, uncommon, small, small population, not abundant	
2	Fairly abundant, fairly common, moderately abundant	
3	Abundant, common, large, large populations	Has to be assigned first to allow for replacement in cases of "not common" or similar combinations
4	Very abundant, very common	

IUCN red list population description get scraped for keywords. All keywords were saved for each species. Keywords were categorized and species assigned according to their keywords In cases of more than one possible assignment, the species gets assigned to the highest category to account for descriptions of species which are uncommon in xyz, but have large popuations in suitable habitats.

Data imputation

Multivariate imputation by chained equations creates multiple imputed datasets depending on different sampling sets of the other available variables, and takes the mean of these imputed datasets as finally imputed value. In multiple steps, plausible values are being drawn from a conditional density distribution that is modelled for each incomplete variable. Imputed variables subsequently enter the next imputation step. For full details on the algorithm behaviour see Van Buuren & Groothuis-Oudshoorn (2011). We chose this data imputation method because it allows for chosing different imputation models for each variable types, as the body size is a continuous numeric variable, the abundance however is implemented as a factor, coded as integer. The mice package allows for this distinction.

The goal of our data imputation was to be able to use the maximum number of occurrences while adding a minimum of noise in the data. The potential influences of data imputation is difficult to analyse, as the missing data cannot be recovered. Hunt (2017) however show, by randomly deleting variables from a complete dataset and completing it again with imputation, that using various multiple data imputation methods achieve around 80% or higher correct classifications in datasets with two or three different classes. This percentage was independant of the proportion of missing values in the data (10 to 50%). We thus assume our imputation method to be comparably effective.

Technical details

The R Code for the data imputation can be found in *Tietje_Roedel_2017_model_building_and_pred*

We used the mice::mice() function with settings maxit=20, m=50 and methods "rf" for body size and "polr" for abundance categories. For comparison with an alternative data imputation method, we are also imputed data using the rf::rfImpute() function. Following Van Buuren & Groothuis-Oudshoorn (2011) we ran diagnostic checks on the imputed data by checking for convergence of the imputing algorithms, plausibility of imputed data, and occurrences of impossible data. The results are shown in the following figures.

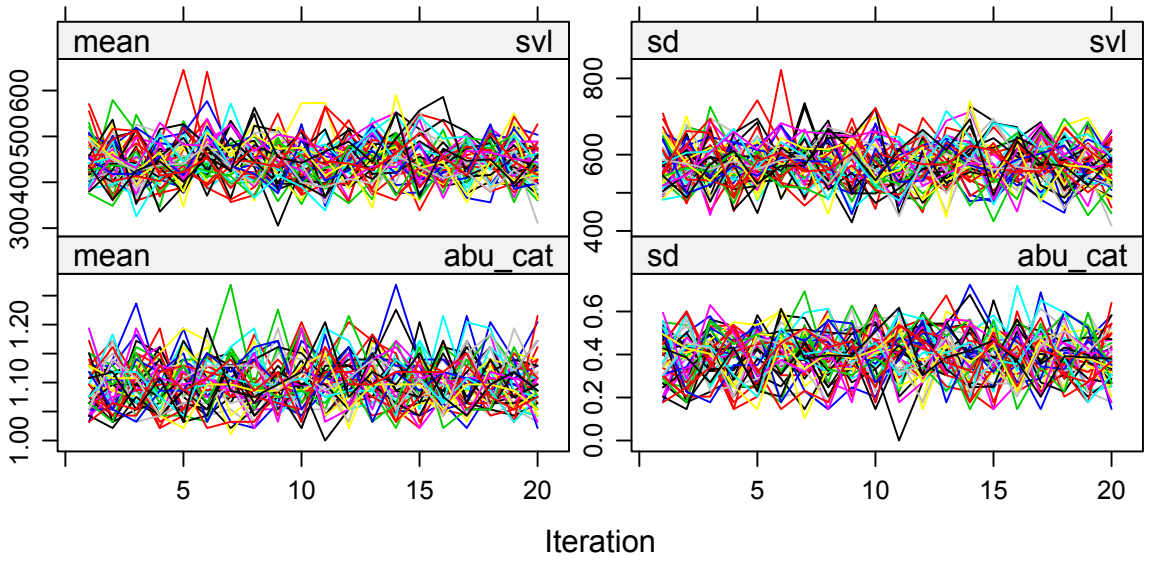

Figure S4: Convergence of the MICE algorithm for body size and abundance. Plotted are the mean and standard deviations of the imputed values per iteration. svl = body size, abu_cat = abundance. For variable explanation see Tab. S1.

Plausibility of the imputed values was checked by comparing density plots of imputed and observed values.

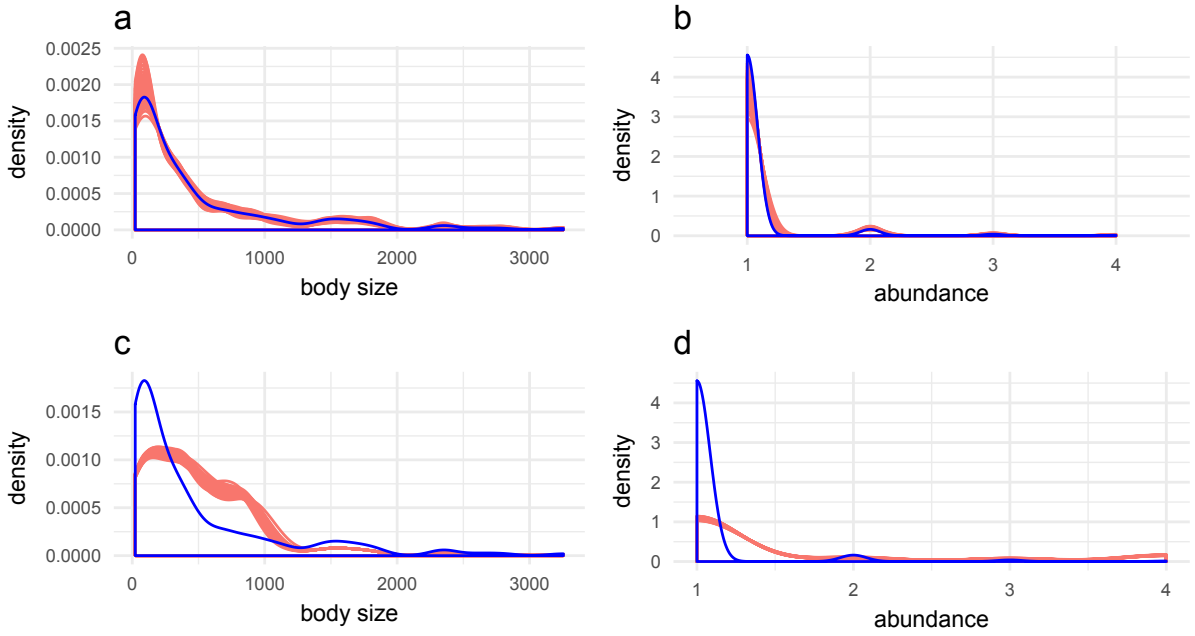

Figure S5: Kernel density estimates of the observed data (blue) and the m=50 densities per variable calculated from the imputed data (red lines) using the MICE algorithm (panel a) and b) and the randomForest algorithm (c and d).

Imputed and observed values for body size and abundance in each group are shown in the following figures.

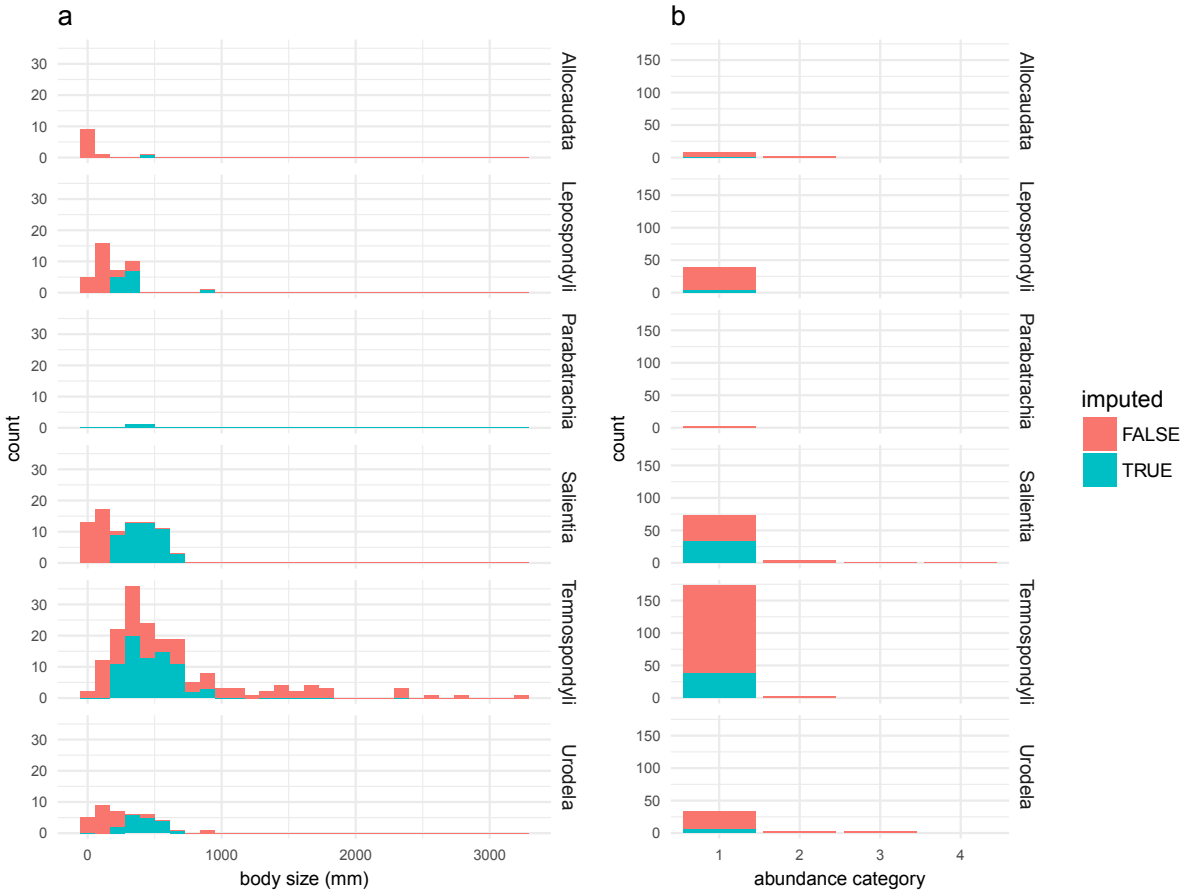

Figure S6: Imputated values in each taxonomic group. Blue color indicates the imputed values. Panel a) shows histograms of the body size in mm, panel b) shows the abundance category data.

Table S4: Percent imputed values in each taxonomic group and variable.

	body size	abundance
Allocaudata	0.09	0.09
Lepospondyli	0.33	0.13
Parabatrachia	1.00	0.00
Salientia	0.61	0.42
Temnospondyli	0.43	0.22
Urodela	0.46	0.18
Total	0.47	0.26

Imputation summary

Of all occurrences, 53% had some missing data, while only 19% of occurrences were missing both body size and abundance measure. Imputed values for the two incomplete variables body size and abundance made up 26% and 47% of those two variables, respectively. Those proportions varied between the taxonomic groups, with Allocaudata showing the highest and Salientia the lowest proportions of imputed data (Fig. S6, Tab. S4). The quality control showed that density plots of observed and imputed values were highly comparable (Fig

S5 a and b). Further, we did not find any impossible values (like negative body size) in the imputed data. Therefore, the quality check suggests that the imputed data is likely reasonable (Van Buuren & Groothuis-Oudshoorn 2011).

The imputation increased the sample size (number of occurrences we were able to ues in the model) by 62%, which added to the stability of the model. Additionally, the imputed values were restricted to the variables abundance and body size, which were of lower importance for the final model. Therefore the influence of uncertainties introduced by data imputation on the outcome of the model should have been rather small. This was also reflected by the minor differences between the GBM fitted on imputed and unimputed data (Fig. S19, S20, Tab. S5).

Model selection

Due to the nature of the dataset (high skewness and kurtosis of the variables) we applied three different models to our data to connect traits with the duration of species: Generalized additive model (GAM), randomForest (rF), and Generalized boosted model (GBM).

The process of selecting the best parameter settings for the models as done in the caret package (Kuhn *et al.* 2017) is illustrated in the following Fig. S7:

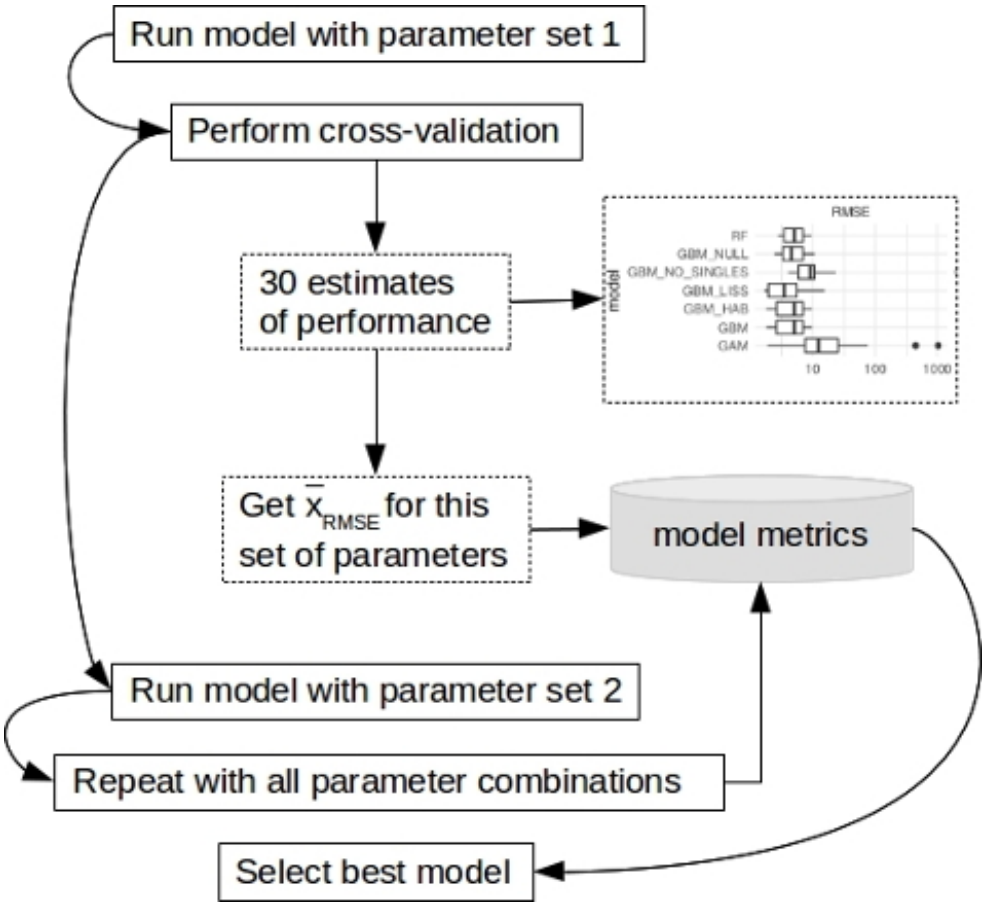

Figure S7: Model selection process, depicted as flowchart. Each model type (GAM, rF and GBM) was fitted using cross-validation. The model performance metrics were collected for each parameter set, finding the best parameter setting for the model.

Generalized additive model

Before adjusting the model parameters, we checked if logging the skewed variables in the extinct species data results in a better model performance (using default settings). While logging positively influenced the GAM fitted to the original dataset including missing values (which simply removes all cases with missing data), it only slighty influenced the GAM fitted on the imputed dataset. We therefore decided to stick to the unlogged, imputed dataset for further modelling.

A generalized additive model was fitted to the extinct species dataset using the caret::train() function. We used 3 separate 10-fold cross validations, meaning the extinct species-dataset was randomly split into 10 equal sized subsamples, from which 9 subsamples were used as

training data and one was retained as validation data for testing the model. This process was repeated until each of the subsamples was used once as validation set. The procedure was repeated 3 times.

Output 1: Console output fitting the gam using caret::train() function.

```
## Generalized Additive Model using Splines
##
## 354 samples
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 319, 318, 319, 319, 319, 318, ...
## Resampling results across tuning parameters:
##
##   select  RMSE      Rsquared   MAE
##   FALSE   62.42695  0.2980428  12.817146
##    TRUE   21.95516  0.2718881   5.682457
##
## Tuning parameter 'method' was held constant at a value of GCV.Cp
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were select = TRUE and method = GCV.Cp.
```

Output 2: Console output for the final gam.

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## .outcome ~ s(lat_range) + s(gcd) + s(min_lat) + s(mean_lat) +
##     s(svl)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.7994     0.1917   9.384   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                  edf Ref.df      F  p-value
## s(lat_range) 7.9139      9  8.412 6.88e-13 ***
## s(gcd)       7.7328      9 12.141  < 2e-16 ***
## s(min_lat)   7.7675      9 41.521  < 2e-16 ***
## s(mean_lat)  7.6842      9 41.645  < 2e-16 ***
## s(svl)       0.8119      9  0.124     0.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.662   Deviance explained = 69.3%
## GCV =  14.35  Scale est. = 13.016    n = 354
```

Random Forest

A Random Forest model was fitted to the extinct species dataset using the caret::train() function. We used 3 separate 10-fold cross validations, meaning the extinct species-dataset was randomly split into 10 equal sized subsamples, from which 9 subsamples were used as training data and one was retained as validation data for testing the model. This process was repeated until each of the subsamples was used once as validation set. The procedure was repeated 3 times.

Output 3: Console output fitting the rF using caret::train() function.

```
## Random Forest
##
## 354 samples
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 319, 318, 319, 319, 319, 318, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##   2     5.198038  0.2492967  2.402612
##   3     5.260152  0.2498780  2.348284
##   4     5.315078  0.2470060  2.322943
##   5     5.366857  0.2398260  2.332688
##   6     5.393832  0.2365635  2.321819
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 2.
```

Output 4: Console output for the final rF.

```
##
## Call:
##  randomForest(x = x, y = y, mtry = param$mtry, importance = TRUE,      verb
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 33.52467
##                     % Var explained: 12.73
```

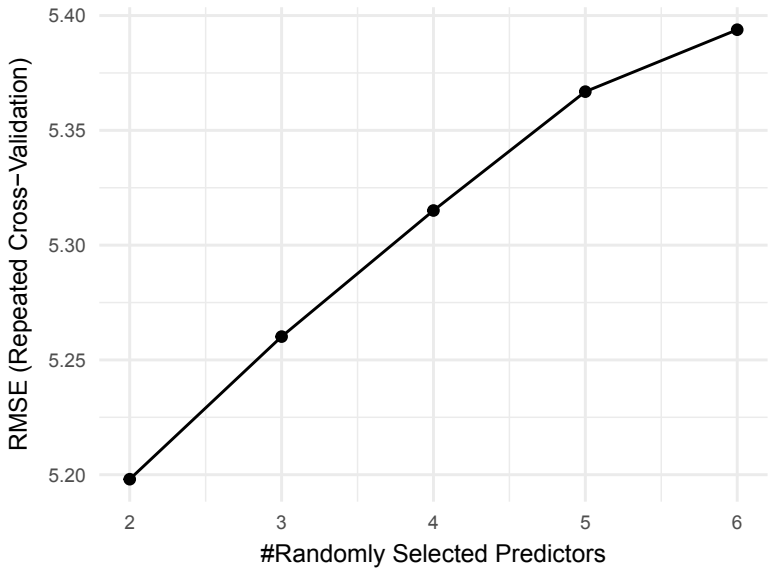

Figure S8: Root mean squared error (RMSE) for each cross validation set of different parameters mtry (number of predictor to chose from at each split in the random Forest).

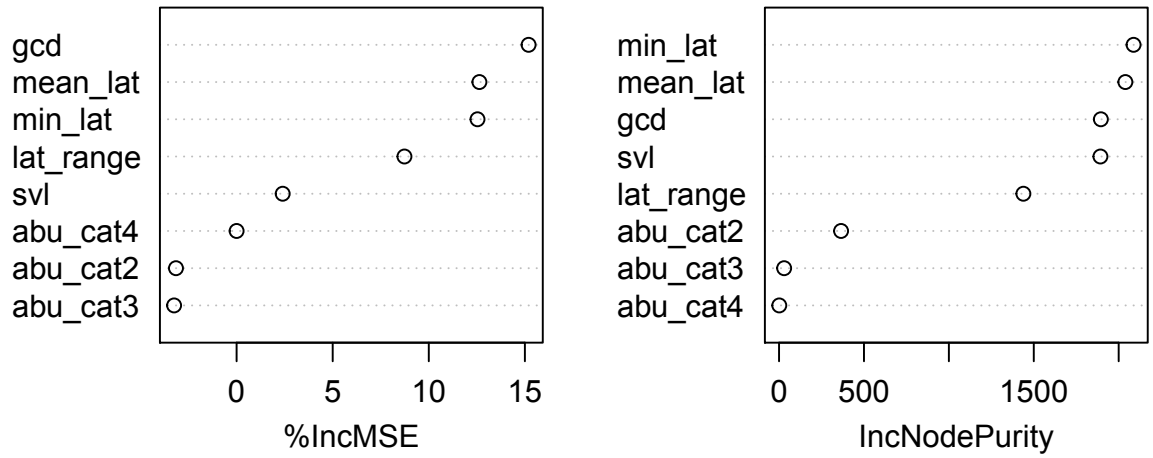

Figure S9: Variable importance measures for the final randomForest. Left plot: Difference in percent between the prediction error (mean squared error (MSE)) of the model and the prediction error after permuting each variable while holding all other data constant (%IncMSE) Differences are normalized by the standard deviation of the differences. Right plot: The total decrease in node impurity from splitting on the variable, averaged over all trees, measured by residual sum of squares (IncNodePurity). Variable names on the y-axis are the variables as used in the model and explained in Table S1.

Generalized boosted model

A generalized boosted model (GBM) was fitted to the extinct species dataset using the caret::train() function. We used the same cross-validation procedure to obtain the optimal tuning parameters as for the GAM and rF.

Output 5: Output fitting the GBM using caret::train() function. For full output run the R script.

```
## Stochastic Gradient Boosting
##
## 354 samples
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 318, 319, 319, 319, 318, 319, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  n.minobsinnode  n.trees  RMSE      Rsquared
##   0.001      1                   5               50      5.516714  0.271712
##   0.001      1                   5              100      5.472123  0.283205
##   0.001      1                   5              150      5.428993  0.283511
##   0.001      1                   5              200      5.391019  0.282354
##   0.001      1                   5              250      5.356412  0.278576
##   0.001      1                   5              300      5.325467  0.277695
##   0.001      1                   5              350      5.296146  0.278573
##   MAE
##   3.052148
##   3.021109
##   2.990499
##   2.962987
##   2.936768
##   2.911829
##   2.887404
##  [ reached getOption("max.print") -- omitted 263 rows ]
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 200, interaction.depth =
##  1, shrinkage = 0.01 and n.minobsinnode = 5.
```

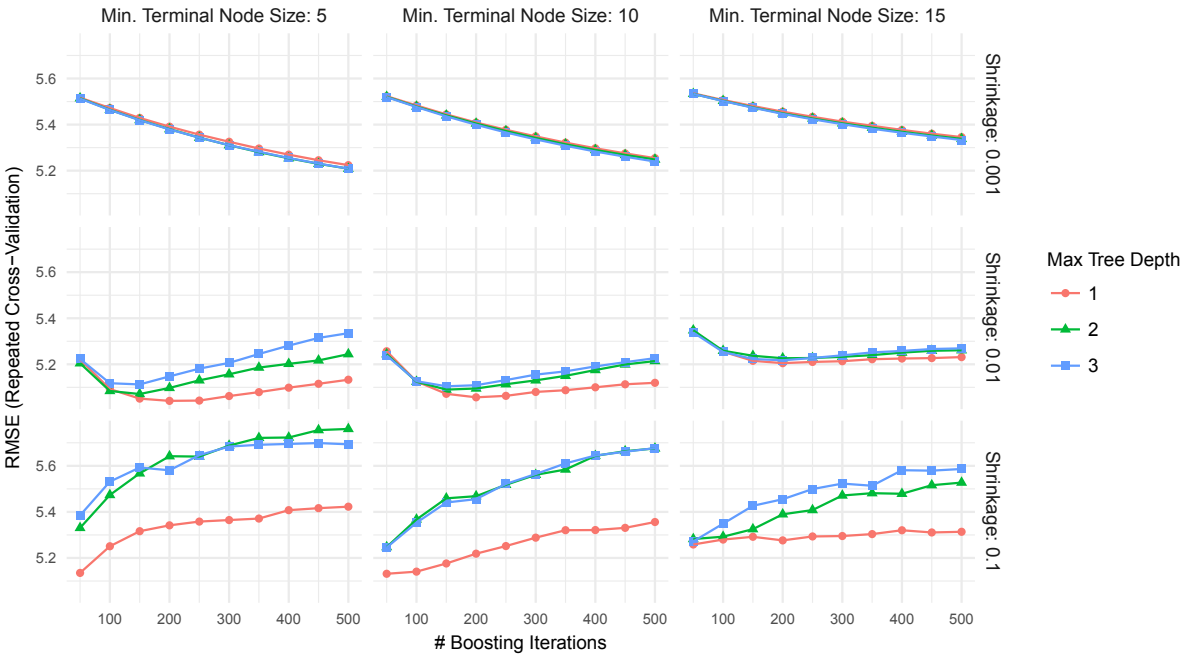

Figure S10: Relationship between the Root-mean-squared error (RMSE) of the GBM as estimate of performance and the tuning parameters: minimum terminal node size (vertical panels), shrinkage (horizontial panels) and maximum tree depth (legend). Minimum terminal node size defines the minimum number of observations in the trees terminal nodes, shrinkage is the learning rate of the model, maximum tree depth is the interaction depth for variables, with 1 being an additive model.

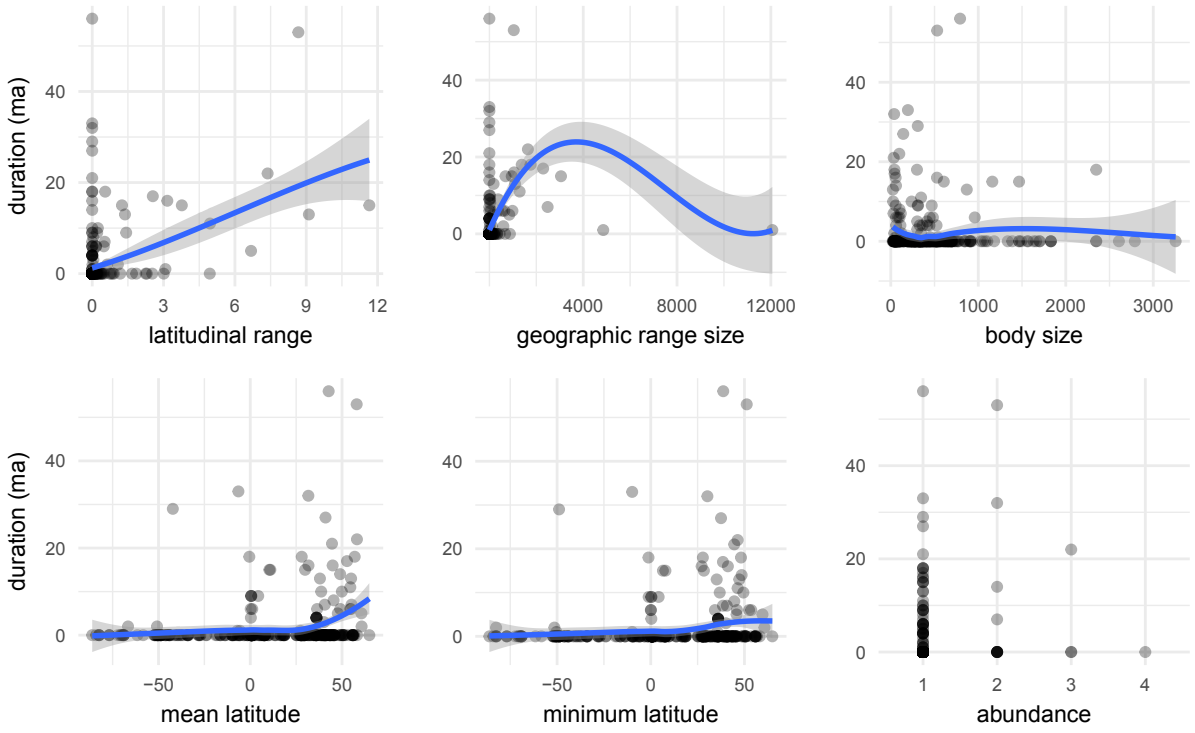

Figure S12: Scatterplots of each prediction variable with the response variable for the fossil species data. Variables are described in detail in Table S1.

Bias

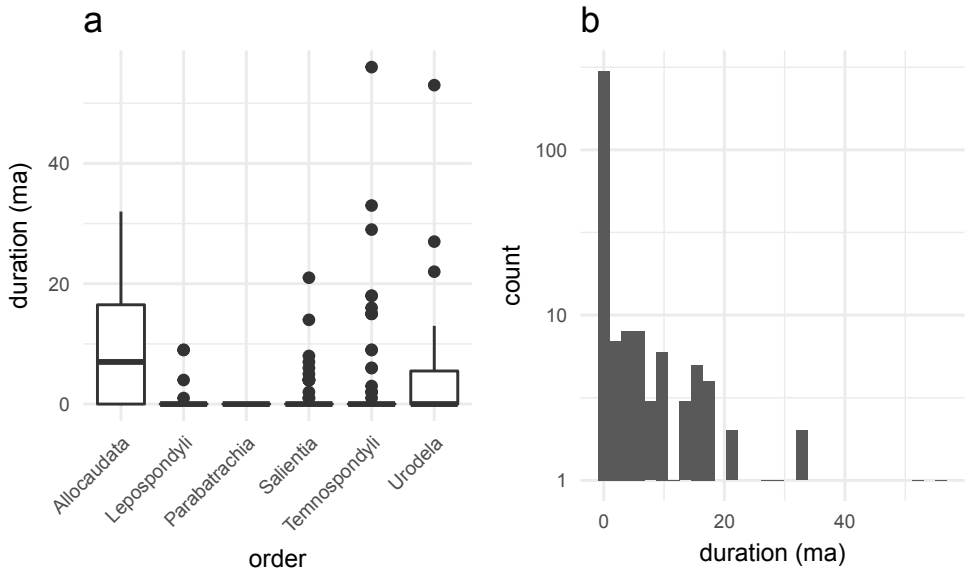

Figure S13: Durations of fossil species as histogram in the main taxonomic groups (a) and as histogram (b). Number of species per group are: Allocaudata (11), Lepospondyli (39), Parabatrachia (2), Salientia (80), Tenmospondyli (175), Urodela (39).

There is a slight taxonomic bias on the duration with Allocaudata having longer durations than Lepospondlyi, Salientia and Temnospondyli. Temnospondyli show shorter duations than Urodela. Allocaudata and Urodela seem to have slightly longer durations on average.

Output 6: Pairwise comparison output from comparing durations between different taxonomic groups.

```
##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  extinct.raw$ma_range and extinct.raw$order
##
##               Allocaudata Lepospondyli Parabatrachia Salientia Temnospondyli
## Lepospondyli  0.0029      -            -             -         -
## Parabatrachia 0.3965      0.7294       -             -         -
## Salientia     0.0037      0.4766       0.6697        -         -
## Temnospondyli 0.0001      0.9889       0.7294        0.2120    -
## Urodela       0.1839      0.0500       0.5270        0.1145    0.0029
##
## P value adjustment method: fdr
```

Adjusting the model for taxonomic differences

To account for differences in model performance between the taxonomic groups, that might be caused by their slightly differing mean stratigraphic ranges, we analysed model performance within each group seperately.

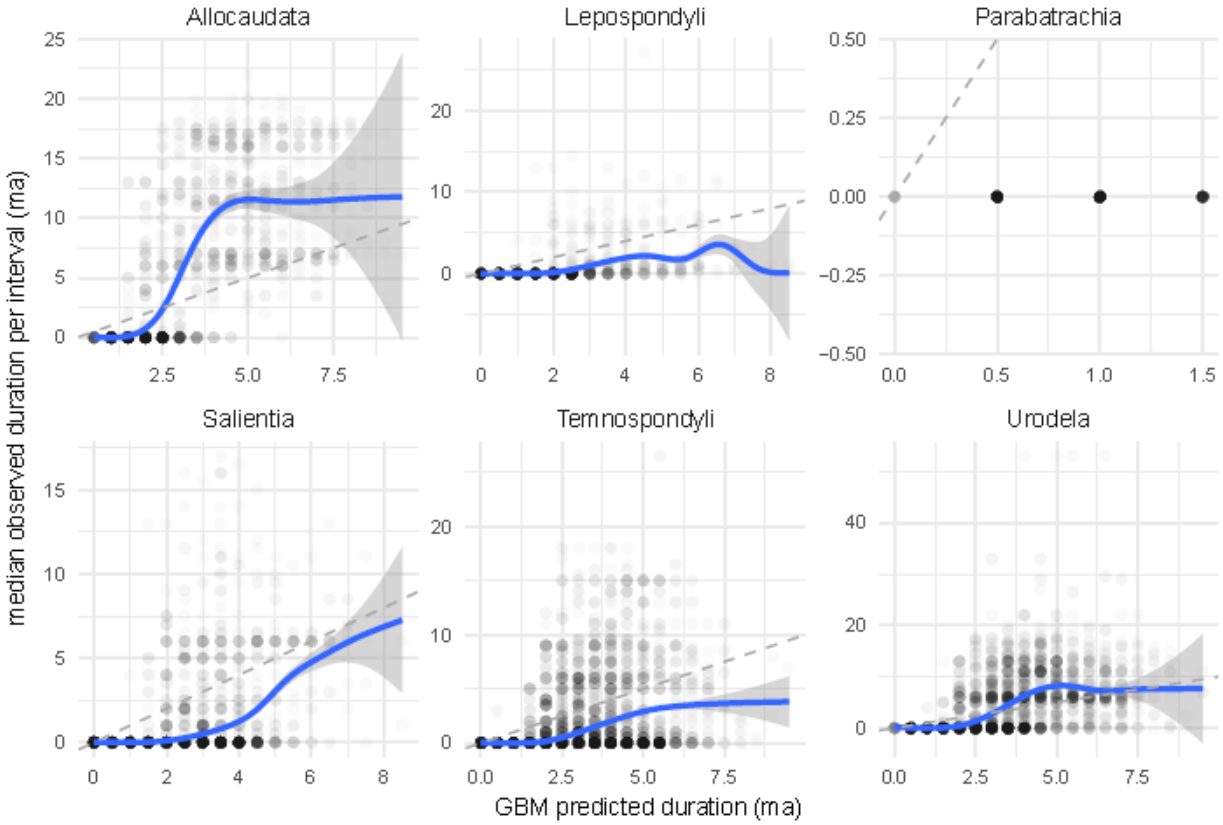

Figure S14: Predicted species durations plotted against observed durations for each taxonomic group. Results are for n=500 repetitions in which the data was randomly split into calibration and test dataset. The blue curve is a generalized additive model with default settings used in geom_smooth() of the ggplot2 package, the grey areas are 95% confidence intervals for the gam. The dashed line is a line through the origin with a slope of 1.

Within Lissamphibia, Salientia seem to be mostly underpredicted in the model, whereas Urodela durations are switching from being underpredicted for short durations to being overpredicted for longer durations. These differences between observed and predicted duration were corrected for in the GBM-CORR predictions (Fig. S20).

Lissamphibia model

To control for potential further influence of these minor differences, we fitted a GBM to the extinct species dataset, using lissamphibian species only (Salientia, Urodela and Parabatrachia) Results are comparable to the GBM on all data in terms of variable importance (Fig. S15) and error estimates (Fig. S19).

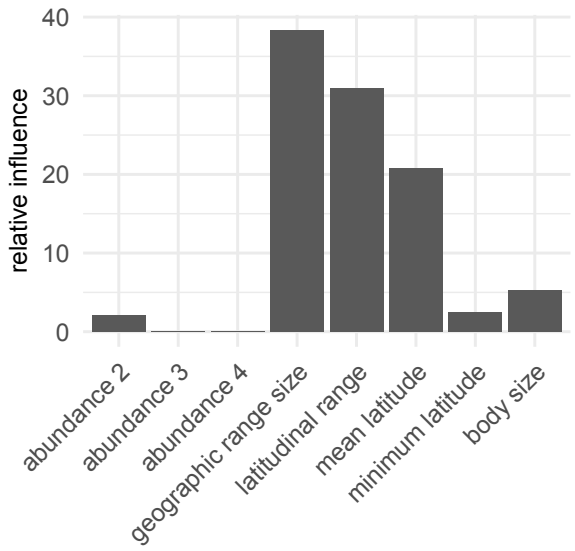

Figure S15: Relative influence of each variable for the final Lissamphibia GBM. The relative influence is an importance measure for the influence of each predictor variable in the model Values are scaled to match 100%. Variables are described in detail in Table S1.

No singleton model

Removing all taxa who appear in one single chronostratigraphic stage is a common praxis in paleontological quantitative data analysis and supposed to reduce the inclusion of false single-interval species due to bad conversation. Although we doubt that the removal of this potential bias towards short durations outweighs the bias introduced by dramatic dataset reduction, accompanied by massive diversity loss, we fitted a model to a subset of our data This subset only includes species which have a duration length greater than 0, meaning they were found in at least two chronostratigraphic stages.

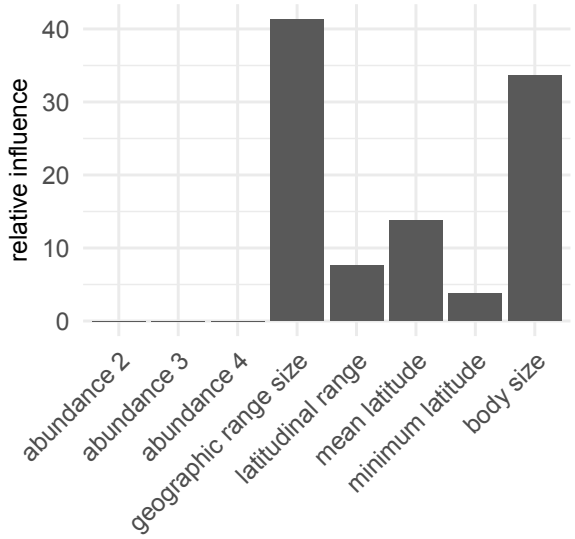

Figure S16: Relative influence of each variable for the no-single-interval species GBM. The relative influence is an importance measure for the influence of each predictor variable in the model. Values are scaled to match 100%. Variables are described in detail in Table S1.

Null model

To test the Null hypothesis that there is no connection between traits and the survival length, and therefore extinction risk, of species, we created a null model by fitting the GBM to a randomized dataset. The only variable which was randomized was the duration, therefore the other trait combinations stayed as they were to avoid having biological unmeaningful combinations.

Output 7: Output of the null GBM.

```
## Stochastic Gradient Boosting
##
## 354 samples
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 318, 319, 319, 319, 318, 319, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  n.minobsinnode  n.trees  RMSE      Rsquared
##   0.001      1                   5               50      5.654018  0.021974
##   0.001      1                   5              100      5.655204  0.020665
##   0.001      1                   5              150      5.656760  0.017016
##   0.001      1                   5              200      5.658972  0.017744
##   0.001      1                   5              250      5.660760  0.017194
##   0.001      1                   5              300      5.662761  0.016984
##   0.001      1                   5              350      5.665063  0.017423
##   MAE
##   3.084217
##   3.084275
##   3.084732
##   3.085677
##   3.086048
##   3.086237
##   3.086703
##  [ reached getOption("max.print") -- omitted 263 rows ]
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 50, interaction.depth =
##  1, shrinkage = 0.001 and n.minobsinnode = 15.
```

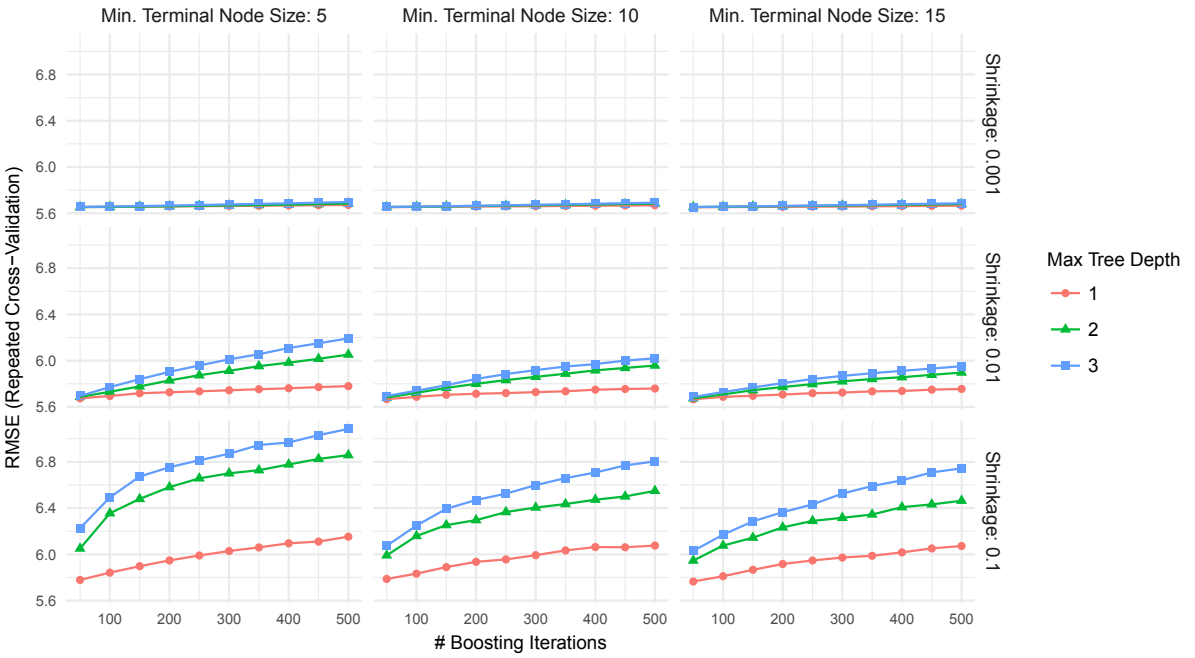

Figure S17: Relationship between the Root-mean-squared error (RMSE) of the GBM as estimate of performance and the tuning parameters: minimum terminal node size (vertical panels), shrinkage (horizontial panels) and maximum tree depth (legend). Minimum terminal node size defines the minimum number of observations in the trees terminal nodes, shrinkage is the learning rate of the model, maximum tree depth is the interaction depth for variables, with 1 being an additive model.

Null model bootstrap

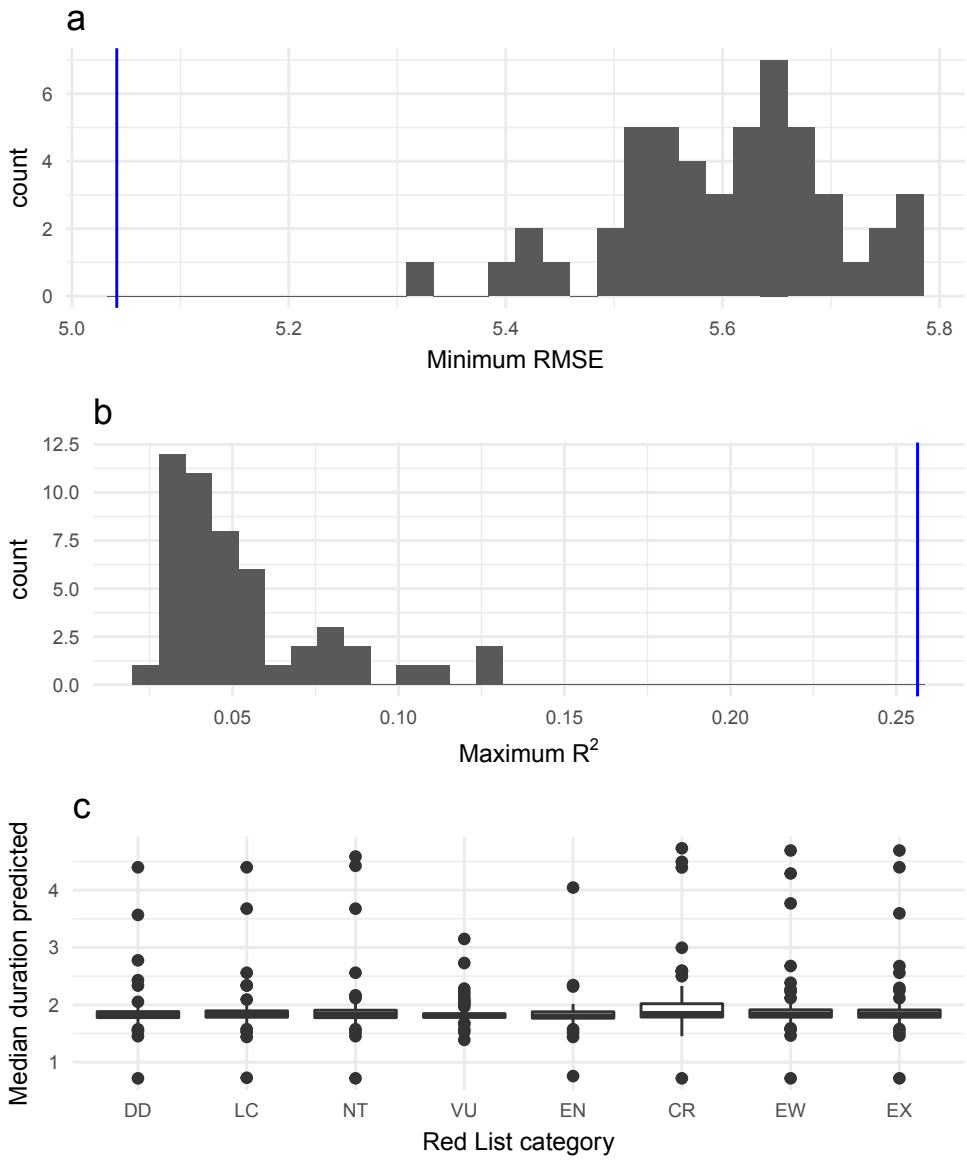

Figure S18: Null GBM bootstrap summaries. A) shows the RMSE values of all 50 null models, b) shows the maximum R^2 in each null model fitting, c) shows the predicted medium durations for the IUCN Red List categories of all 50 null models. The blue lines depict the corresponding values for RMSE and R^2 from the full GBM.

Model comparison

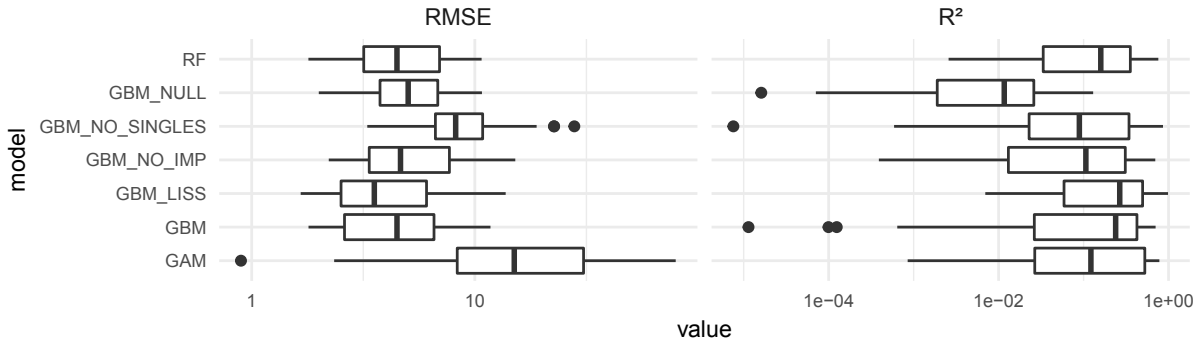

Figure S19: Model performance distributions for the final model cross validation. Each model configuration is tested in a 3 times 10-fold cross validation procedure, the results are displayed here. Sample size is n=30 for each distribution, in accordance with the 3 times 10-fold cross validation procedure. The final model is chosen based on the mean RMSE values of these distribution. RF = randomForest, GBM_NULL = GBM Null model, GBM_NO_SINGLES = GBM without single interval species, GBM_NO_IMP = GBM on data without imputation, GBM_LISS = GBM on Lissamphibia only, GBM = Generalized boosted model, GAM = Generalized additive model.

Table S5: Model comparison. The average RMSE (Root mean squared error) and R^2 including their standard deviations (SD) associated with the optimal tuning parameters across the resamples from cross validation. N=30 for each model, like in Fig. S19.

	RMSE	R^2	RMSE SD	R^2 SD
GBM	5.04	2.79	0.26	0.23
GBM no imputation	5.93	3.65	0.22	0.24
GBM Lissamphibia	4.86	3.48	0.32	0.30
GBM Null model	5.65	2.61	0.02	0.03
random Forest	5.20	2.58	0.25	0.25
GBM no single interval species	10.02	5.99	0.22	0.27
GAM	21.96	20.20	0.27	0.27

In terms of RMSE, the only model being significantly different from the others was the model excluding all single interval taxa, which was not surprising given the much lower sample size. The R^2 value of the null model differed significantly from all other models.

Output 8: Model comparison t-Test.

```
##
## Call:
## summary.diff.resamples(object = difValues, round = 2)
##
## p-value adjustment: fdr
## Upper diagonal: estimates of the difference
## Lower diagonal: p-value for H0: difference = 0
##
## MAE
##                GBM       GBM_NO_IMP GBM_LISS  RF        GBM_NULL  GAM
## GBM                      -1.0920    -0.2444    0.1701   -0.5105   -3.1098
## GBM_NO_IMP     0.001857              0.8476    1.2621    0.5815   -2.0178
## GBM_LISS       0.426580  0.012074              0.4144   -0.2661   -2.8654
## RF             0.412882  7.525e-05  0.145852           -0.6806   -3.2798
## GBM_NULL       0.011968  0.071834   0.334138  0.003172           -2.5993
## GAM            0.001511  0.020294   0.001857  0.000273  0.004350
## GBM_NO_SINGLES 1.389e-06 1.001e-05  1.623e-06 6.617e-07 1.924e-06 0.052298
##                GBM_NO_SINGLES
## GBM            -5.2262
## GBM_NO_IMP     -4.1341
## GBM_LISS       -4.9818
## RF             -5.3962
## GBM_NULL       -4.7157
## GAM            -2.1164
## GBM_NO_SINGLES
##
## RMSE
##                GBM       GBM_NO_IMP GBM_LISS  RF        GBM_NULL  GAM
## GBM                       -0.8913     0.1849   -0.1567   -0.6110  -16.9138
## GBM_NO_IMP     0.4704483              1.0762    0.7346    0.2803  -16.0225
## GBM_LISS       0.8471417 0.4486286             -0.3416   -0.7960  -17.0988
## RF             0.8471417 0.4881829  0.7882626           -0.4543  -16.7571
## GBM_NULL       0.5185537 0.8363928  0.4620276 0.6192691           -16.3028
## GAM            0.0007052 0.0007052  0.0007052 0.0007052 0.0007052
## GBM_NO_SINGLES 0.0013219 0.0030928  0.0009961 0.0007052 0.0010864 0.0059675
##                GBM_NO_SINGLES
## GBM             -4.9740
## GBM_NO_IMP      -4.0827
## GBM_LISS        -5.1589
## RF              -4.8172
## GBM_NULL        -4.3629
## GAM             11.9399
## GBM_NO_SINGLES
##
## Rsquared
##                GBM       GBM_NO_IMP GBM_LISS  RF        GBM_NULL  GAM
## GBM                       0.039199  -0.065800  0.012810  0.233964 -0.009782
## GBM_NO_IMP     0.8273996            -0.104999 -0.027066  0.194765 -0.049657
```

```
## GBM_LISS       0.7666365 0.4836255            0.064931  0.299764  0.042340
## RF             0.9030371 0.8273996  0.8273996           0.226103 -0.022591
## GBM_NULL       7.634e-05 0.0008001  7.634e-05 0.0001940           -0.248694
## GAM            0.9260584 0.8273996  0.8273996 0.8273996 0.0001940
## GBM_NO_SINGLES 0.8273996 0.9260584  0.5143430 0.9030371 0.0014623 0.8273996
##                GBM_NO_SINGLES
## GBM             0.032601
## GBM_NO_IMP     -0.006599
## GBM_LISS        0.098401
## RF              0.019881
## GBM_NULL       -0.201364
## GAM             0.042473
## GBM_NO_SINGLES
```

Predictions

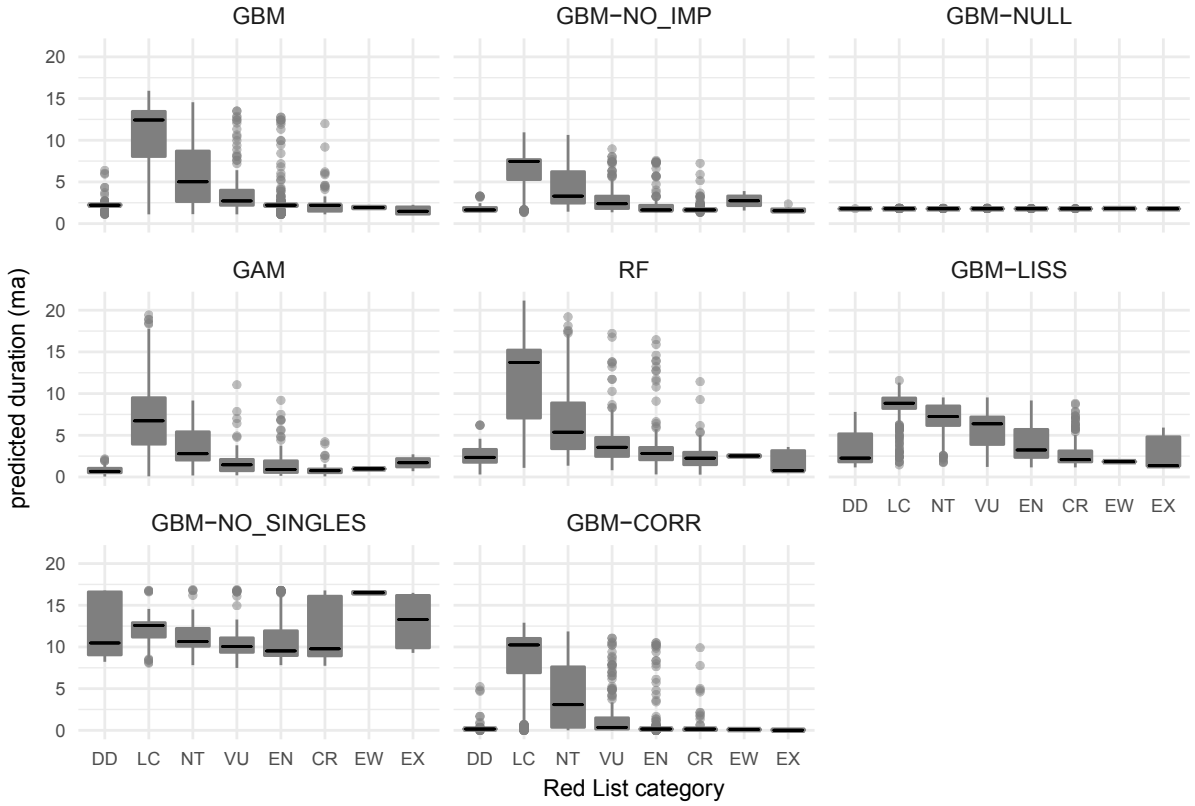

Figure S20: Predicted durations of living species based on all build models. Note that the predicted durations of the GAM have been square-root transformed for better visibility, as the range of durations was much larger than for the other models. GBM = Generalized boosted model, GBM-NO_IMP = GBM on data without imputation, GBM_NULL = GBM Null model, GAM = Generalized additive model, RF = randomForest, GBM_LISS = GBM on Lissamphibia only, GBM_NO_SINGLES = GBM without single interval species, GBM-CORR = GBM with taxonomic bias correction (Fig. S14).

Statistical analysis of the prediction results

The pairwise comparisons of predicted values per group of IUCN Red List extinction risk category are depicted in the following output files.

Output 9: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the gam.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.gam and res$Red.List.status
## Kruskal-Wallis chi-squared = 368.34, df = 7, p-value < 2.2e-16

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  as.numeric(res$predict.gam) and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC 1.3e-15 -       -       -       -       -       -
## NT 0.01767 5.1e-12 -       -       -       -       -
## VU 0.55831 < 2e-16 0.00079 -       -       -       -
## EN 0.00679 < 2e-16 0.00014 0.54427 -       -       -
## CR 0.04830 < 2e-16 0.00066 0.85147 0.27342 -       -
## EW 0.96205 0.30050 0.84154 0.85147 0.55831 0.55831 -
## EX 0.87223 0.04830 0.61962 0.88826 0.76062 0.85147 1.00000
##
## P value adjustment method: fdr
```

Output 10: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the rF.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.rf and res$Red.List.status
## Kruskal-Wallis chi-squared = 746.55, df = 7, p-value < 2.2e-16

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  res$predict.rf and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC < 2e-16 -       -       -       -       -       -
## NT 2.6e-14 3.6e-16 -       -       -       -       -
## VU 2.5e-07 < 2e-16 7.4e-07 -       -       -       -
## EN 0.02173 < 2e-16 9.2e-15 4.7e-05 -       -       -
## CR 0.35076 < 2e-16 < 2e-16 3.5e-13 1.6e-05 -       -
## EW 0.87025 0.03271 0.10943 0.31451 0.62282 0.63518 -
## EX 0.31451 0.00048 0.00504 0.02740 0.13439 0.35076 0.87025
```

```
##
## P value adjustment method: fdr
```

Output 11: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the GBM.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.gbm1 and res$Red.List.status
## Kruskal-Wallis chi-squared = 734.79, df = 7, p-value < 2.2e-16

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  res$predict.gbm1 and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC < 2e-16 -       -       -       -       -       -
## NT 2.0e-10 < 2e-16 -       -       -       -       -
## VU 7.9e-05 < 2e-16 9.7e-06 -       -       -       -
## EN 0.34523 < 2e-16 1.4e-13 1.9e-05 -       -       -
## CR 0.17458 < 2e-16 < 2e-16 1.8e-09 0.00752 -       -
## EW 0.32261 0.03365 0.09321 0.18873 0.32063 0.62102 -
## EX 0.05961 0.00043 0.00261 0.01152 0.04602 0.12650 0.59259
##
## P value adjustment method: fdr
```

Output 12: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the GBM on Lissamphibia.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.gbm.liss and res$Red.List.status
## Kruskal-Wallis chi-squared = 789.38, df = 7, p-value < 2.2e-16

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  res$predict.gbm.liss and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC < 2e-16 -       -       -       -       -       -
## NT < 2e-16 < 2e-16 -       -       -       -       -
## VU 7.1e-12 < 2e-16 6.9e-06 -       -       -       -
## EN 0.00077 < 2e-16 < 2e-16 6.1e-11 -       -       -
## CR 0.43395 < 2e-16 < 2e-16 < 2e-16 2.7e-08 -       -
## EW 0.46541 0.02305 0.03351 0.04943 0.08848 0.46541 -
## EX 0.31384 0.00036 0.00205 0.01338 0.17335 0.43395 0.85714
##
## P value adjustment method: fdr
```

Output 13: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the GBM on the subset without single-interval species.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.gbm.nosingles and res$Red.List.status
## Kruskal-Wallis chi-squared = 194.88, df = 7, p-value < 2.2e-16

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  res$predict.gbm.nosingles and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC 0.22917 -       -       -       -       -       -
## NT 0.91852 2.8e-10 -       -       -       -       -
## VU 0.21681 < 2e-16 0.00017 -       -       -       -
## EN 0.04342 < 2e-16 1.2e-05 0.04342 -       -       -
## CR 0.30690 0.00057 0.15248 0.83872 0.22227 -       -
## EW 0.21831 0.04679 0.05977 0.05977 0.09672 0.16662 -
## EX 0.70278 0.49778 0.47507 0.21831 0.17457 0.33949 0.25397
##
## P value adjustment method: fdr
```

Output 14: Kruskal-Wallis rank sum test and Pairwise Wilcoxon rank sum test console outputs for the null model.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  res$predict.gbm.null and res$Red.List.status
## Kruskal-Wallis chi-squared = 72.278, df = 7, p-value = 5.113e-13

##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  res$predict.gbm.null and res$Red.List.status
##
##    DD      LC      NT      VU      EN      CR      EW
## LC 0.01326 -       -       -       -       -       -
## NT 0.17509 0.79977 -       -       -       -       -
## VU 0.88359 0.00034 0.09811 -       -       -       -
## EN 0.90963 3.4e-05 0.08966 0.90963 -       -       -
## CR 0.08966 3.3e-11 0.00068 0.09811 0.04727 -       -
## EW 0.11502 0.21762 0.36991 0.17509 0.13233 0.09086 -
## EX 0.88359 0.74059 0.88359 0.88359 0.88359 0.47332 0.32105
##
## P value adjustment method: fdr
```

Potentially misclassified species

Using our model to identify potential misclassifications in the IUCN Red List assignments could help focusing the limited conservation actions to the right species. However, a classification is proving difficult as our model is a numerical model, not a classfication model. Therefore, a species that is indeed falsely classified within the Red List will likely show up with an unusually long or short predicted duration in this category, as we predict durations, but not cateogries.

While there are more ways to identify potential misclassification, we chose the simplest one and defined a misclassfication as a species whos duration plots outside the whiskers of the predicted duration boxplots in each catergory (Fig. 3); that is durations either larger than the third quartile + 1.5 * IQR, or shorter than the first quartile - 1.5 * IQR. The following table (Tab S6) summarizes these potential misclassifications for each Red List status.

Table S6: Number of potentially misclassified species in the IUCN Red List. Misclassifications were defined as statistical outliers, as seen in the prediction boxplots (Fig. 3).

	DD	LC	NT	VU	EN	CR	EW	EX
Longer	13	0	0	19	42	10	0	0
Shorter	15	0	0	0	44	0	0	0
Total	66	740	98	151	192	128	2	5
Longer %	20	0	0	13	22	8	0	0
Shorter %	23	0	0	0	23	0	0	0

References

Harting, J. & Wong, M. (1979). Algorithm AS 136: a k-means clustering algorithm. *J. R. Stat. Soc. C*, 28, 100–108.

Hunt, L.A. (2017). Missing data imputation and its effect on the accuracy of classification. In: *Data science. innovative developments in data analysis and clustering* (eds. Palumbo, F., Montanari, A. & Vichi, M.). Springer Nature, Cham, Switzerland, pp. 3–14.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C. & Engelhardt, A. *et al.* (2017). *caret: Classification and Regression Training*.

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). Multivariate Imputation by Chained Equations. *J. Stat. Softw.*, 45, 1–67.

Xie, Y. (2014). knitr: A Comprehensive Tool for Reproducible Research in {R}. In: *Implementing reproducible computational research* (eds. Stodden, V., Leisch, F. & Peng, R.D.). Chapman; Hall/CRC.

Xie, Y. (2015). *Dynamic Documents with {R} and knitr*. 2nd edn. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y. (2016). *knitr: A General-Purpose Package for Dynamic Report Generation in R*.

Supplementary material: Paper 3

Climate influences the effect of range size on extinction risk in the amphibian fossil record

Supplementary material - Climate influences the effect of range size on extinction risk in the amphibian fossil record

Supplementary figures, tables

Melanie Tietje (Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany)
Mark-Oliver Rödel (Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Germany; Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Germany)
Martin Schobben (Universiteit Utrecht, Utrecht, The Netherlands)

Contents

Intro

This is an R Markdown document which produces the supplement as output. To run the present file you need to run the MS3_5.Rmd file first to create the workspace.RData file, or download that workspace.RData file.

This document contains all supplementary material.

This document is writen with the knitr package (Xie 2014, 2015, 2016) for the R statistical environment R Core Team (2018). The entire analysis can be recreated by running knitr::knit()

The analysis and data files can be accessed in the Git repository: https://github.com/Eryops1/....

We used R version 3.4.4 (2018-03-15) and the following packages:

```
loadedNamespaces()
```

```
##  [1] "Rcpp"         "pillar"       "compiler"     "captioner"
##  [5] "plyr"         "methods"      "utils"        "tools"
##  [9] "grDevices"    "rpart"        "digest"       "evaluate"
## [13] "tibble"       "gtable"       "lattice"      "rlang"
## [17] "Matrix"       "yaml"         "parallel"     "gbm"
```

```
## [21] "proto"        "gridExtra"    "stringr"      "knitr"
## [25] "graphics"     "datasets"     "stats"        "rprojroot"
## [29] "grid"         "nnet"         "reshape"      "mice"
## [33] "base"         "tcltk"        "survival"     "gsubfn"
## [37] "rmarkdown"    "ggplot2"      "magrittr"     "backports"
## [41] "scales"       "htmltools"    "MASS"         "splines"
## [45] "randomForest" "colorspace"   "stringi"      "lazyeval"
## [49] "munsell"
```

Isotope data

Isotope data composition and gradient

The oxygen isotope composition of calcitic fossils is to date the temporally most-resolved and robust climate record for the Phanerozoic. Splicing of low-latitude oxygen isotope data obtained from brachiopod, belemnite and bivalve with planktonic foraminiferal $\delta^{18}O$ for a Phanerozoic reference frame is a best effort to gauge long-term climatic changes, for two reasons; 1) the pre-Cretaceous sediment record consists almost exclusively out of shelf and epeiric sea deposits of shallow water depth, and 2) Veizer & Prokoph (2015) showed that systematic $\delta^{18}O$ offsets between those fossils groups are relatively small comparative to the overall scatter in the dataset. In a similar sense, the influence of preferential 16O storage in icecaps, with a ~1 per mill offset between a glaciated and an ice-free world, would only be marginal in comparison to the overall data-spread of several per mill. This tropical $\delta^{18}O$ record of uniform water depth is therefore sensitive to climate change over a vast area of the Earth. The tropics are, furthermore, sensitive to changes which can have a far-reaching effect on the rest of the globe (Chiang 2009), and thus this record provides an important diagnostic tool to test the effect of climate on amphibian extinction risk.

The meridional $\delta^{18}O$ gradient clearly stands-out from the scatter of the long-term $\delta^{18}O$ record with a several per mill offset between low- and high latitudes during the Early Cretaceous and during much of the Cenozoic. Although meridional energy transport nowadays results in a strong latitudinal temperature gradient, the mechanism(s) of energy transport might have varied through time (Hay & DeConto 1999). For example, it has been postulated that Late Cretaceous meridional seawater overturning was forced by a low-latitude haline-controlled circulation as opposed to a high-latitude deep water formation, which is intimately tied to sea-ice formation in the Cenozoic. In such a uniform warm world, higher levels of humidity in the atmosphere could have transported energy as latent heat in order to maintain the energy balance between low- and high latitudes (Hay & DeConto 1999). This notion is in accordance with the discrepancy between a lower tropical temperature trend and the almost absent meridional temperature gradient during the Late Cretaceous, suggestive of uniform temperatures from pool-to-equator. This observation stipulates the importance of also using meridional $\delta^{18}O$ gradients in assessing the effect of species geographic range size on extinction risk, as long-term tropical $\delta^{18}O$ trends might not in all climatic configurations be as informative in regard to spatial environmental gradients.

The two temperature proxies we used in this study align well with current understanding of climatic trends throughout the Phanerozoic (Prokoph & Veizer 1999). The tropics are sensitive to changes which can have a far-reaching effect on the rest of the globe (Chiang 2009), and thus this record provides an important diagnostic tool to test the effect of climate on biotic

variables. Additionally, we used meridional δ^{18}O gradients in assessing the effect of species geographic range size on extinction risk, as long-term tropical δ^{18}O trends might not in all climatic configurations be as informative in regard to spatial environmental gradients.

Correlation of deep and surface water

Deep and surface water oxygen isotope values are strongly correlated (Spearmans rho= 0.69, p<0.001).

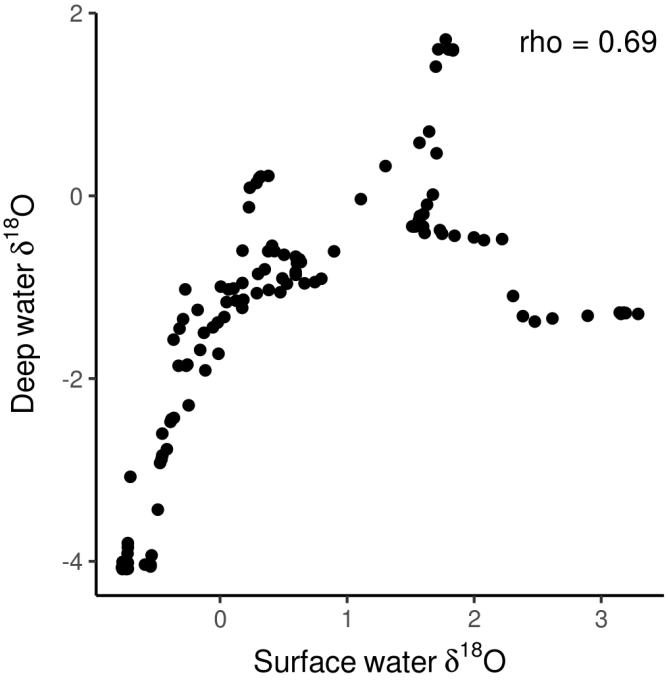

Figure S1: Oxygen isotope values from Prokoph & Veizer 2015, showing deep water oxygen values (classic Zachos curve) and surface values from planctic organisms.

Geographic range size and climate data

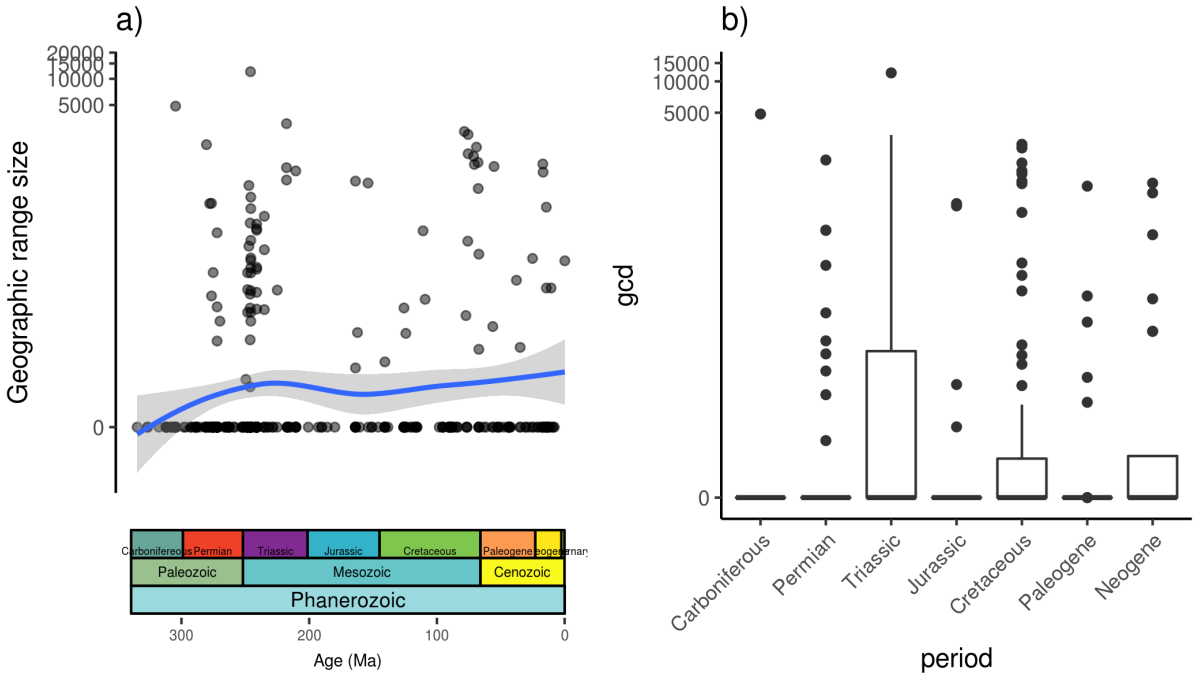

Figure S2: a) Geographic range sizes of 354 amphibian species. Geographic range size was calculated as maximum great circle distance of a species ever achieved over its lifespan. Range size was plotted at the mean age of the corresponding species lifespan. b) Geographic range sizes averaged in each geological Period. There is no significant differences between the Periods.

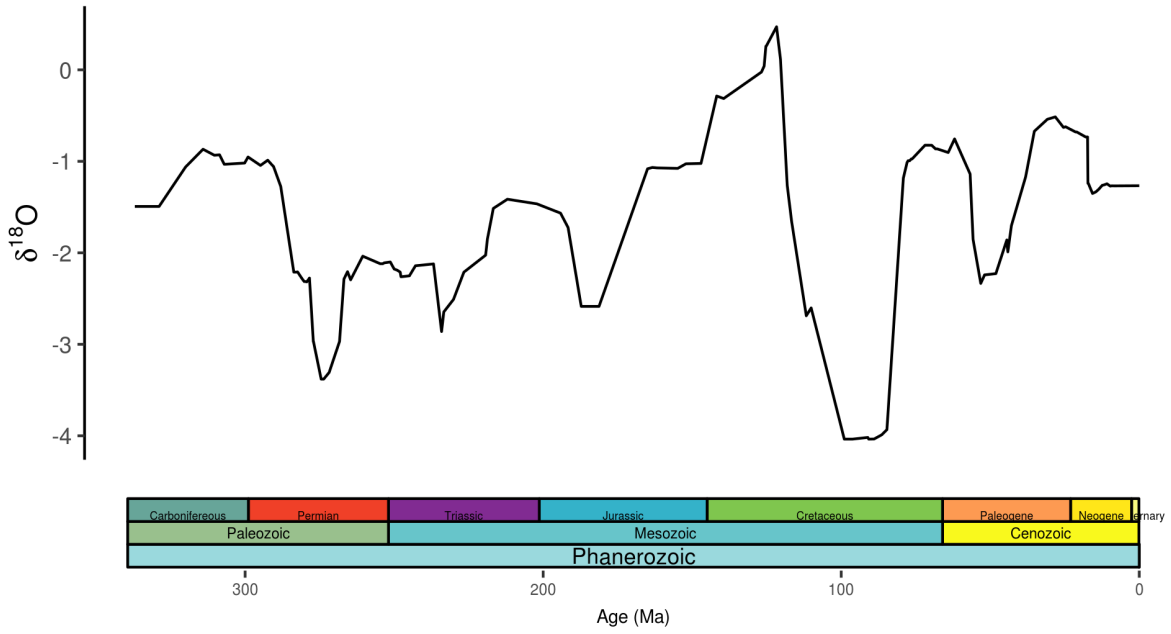

Figure S3: Mean temperature trend for the Phanerozoic depicted as oxgen isotope data from Prokov&Veizer 2015. Values are bootstrapped moving window means of tropical oxygen isotope values.

Geographic range size and latitudinal temperature gradient

Geographic range size was correlated to the temperature gradient (rho= -0.33, p<0.01). This correlation has shown to be stable in cross-validation via bootstrap (figure S3). Cross-validation was performed by repeatedly calculating the correlation on subsets of the data (n=1000, leaving out one third of the data in each subsample).

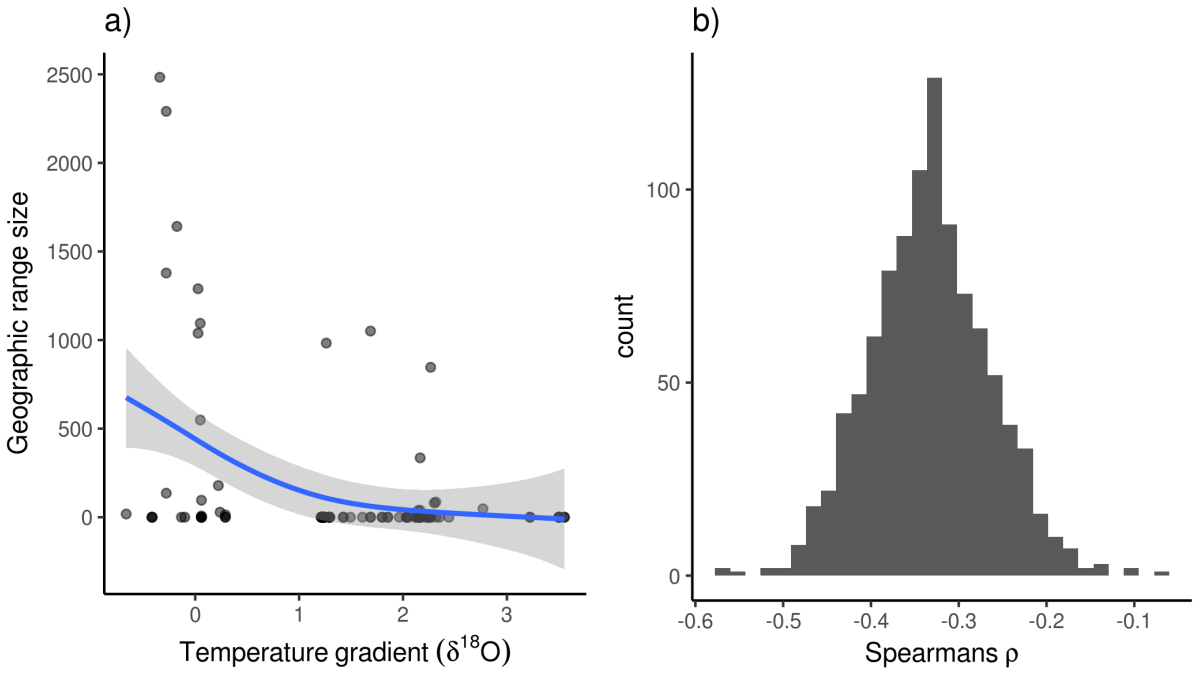

Figure S4: a) Geographic range size of species with the corresponding temperature gradient values at the mean time of the species occurrence. Geographic range sizes are maximum great circle distances, temperature gradient values are differences between high and low paleolatitude oxygen isotope values. Blue line depicts a generalized additive model to show the general trend in the data, grey areas depict the standard error. b) Variation in the correlation of geographic range size and the temperature gradient, assessed with repeated subsampling (leaving out 1/3 of the data in each repetition, n=1000). Correlation was assessed using Spearmans rank correlation.

References

Chiang, J.C. (2009). The Tropics in Paleoclimate. *Annual Review of Earth and Planetary Sciences*, 37, 263–297.

Hay, W.W. & DeConto, R.M. (1999). Comparison of modern and Late Cretaceous meridional energy transport and oceanology. *Special Paper 332: Evolution of the Cretaceous Ocean-Climate System*, 283–300.

Prokoph, A. & Veizer, J. (1999). Trends, cycles and nonstationarities in isotope signals of phanerozoic seawater. *Chemical Geology*, 161, 225–240.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna,

Austria. https://www.R-project.org/.

Veizer, J. & Prokoph, A. (2015). Temperatures and oxygen isotopic composition of Phanerozoic oceans. *Earth-Science Reviews*, 146, 92–104.

Xie, Y. (2014). knitr: A Comprehensive Tool for Reproducible Research in {R}. In: *Implementing reproducible computational research* (eds. Stodden, V., Leisch, F. & Peng, R.D.). Chapman; Hall/CRC.

Xie, Y. (2015). *Dynamic Documents with {R} and knitr*. 2nd edn. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y. (2016). *knitr: A General-Purpose Package for Dynamic Report Generation in R*.

Declaration of independance / Eigenständigkeitserklärung

I hereby certify that I have written the present work independently and have used no other than the specified tools. The parts of the work taken from other works, either verbally or in terms of content, have been identified by corresponding information from the sources. The underlying doctorate regulations are known to me, the work corresponds to the principles of the Humboldt University in Berlin to ensure good scientific practice.

This work did not exist in the same or similar form to any examining authority.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Werken wörtlich oder inhaltlich entnommen sind, wurden durch entsprechende Angaben der Quellen kenntlich gemacht. Die zugrunde liegende Promotionsordnung ist mir bekannt, die Arbeit entspricht den Grundsätzen der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Melanie Tietje, Berlin, den 31. Juli 2018